

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 12:27:29 ; Search time 1258.03 Seconds
(without alignments)
11078.215 Million cell updates/sec

Title: US-09-462-846-1
Perfect score: 945
Sequence: 1 atgcagctgaacctat.....aatgatactctccacatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373928652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_om:*
- 9: gb_pac1:*
- 10: gb_pac2:*
- 11: gb_ph:*
- 12: gb_pi1:*
- 13: gb_pi2:*
- 14: gb_pi3:*
- 15: gb_pi4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_hgo_hum:*
- 20: em_hgo_inv:*
- 21: em_hgo_r0d:*
- 22: em_hgo_hum1:*
- 23: em_hgo_hum2:*
- 24: em_hgo_hum3:*
- 25: em_hgo_hum4:*
- 26: em_hgo_hum5:*
- 27: em_hgo_hum6:*
- 28: em_hgo_hum7:*
- 29: em_hgo_hum8:*
- 30: em_hgo_inv1:*
- 31: em_hgo_inv2:*
- 32: em_hgo_other:*
- 33: em_hgo_r0d:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pac:*

46: em_ph:*

47: em_pi:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_v1:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_v11:*

59: gb_v12:*

60: gb_v13:*

61: gb_v14:*

62: gb_v15:*

63: gb_v16:*

64: gb_v17:*

65: gb_v18:*

66: gb_v19:*

67: gb_v20:*

68: gb_v21:*

69: gb_v22:*

70: gb_v23:*

71: gb_v24:*

72: gb_v25:*

73: gb_v26:*

74: gb_v27:*

75: gb_v28:*

76: gb_v29:*

77: gb_v30:*

78: gb_v31:*

79: gb_v32:*

80: gb_v33:*

81: gb_v34:*

82: gb_v35:*

83: gb_v36:*

84: gb_v37:*

85: gb_v38:*

86: gb_v39:*

87: gb_v40:*

88: gb_v41:*

89: gb_v42:*

90: gb_v43:*

91: gb_v44:*

92: gb_v45:*

93: gb_v46:*

94: gb_v47:*

95: gb_v48:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	945	100.0	945	9	AX006200
2	945	100.0	216750	2	BSDB0007
3	320.4	33.9	36308	2	D88802
4	320.4	33.9	213190	2	BSDB0004
5	319	33.8	945	9	AX006205
6	311.8	33.0	948	2	AX006203
7	311.8	33.0	3962	2	BSDB02562
8	311.8	33.0	4162	2	BACORFX
9	311.8	33.0	212610	2	BSDB0019
10	304	32.2	15012	1	AB011838
11	304	32.2	305153	2	AP001520

12 277.2 29.3 10607 1 AE006310 AE006310 Lactococc
13 265.8 28.1 1379 3 STREPM D16594 S.mutans pm
14 252.2 26.7 49897 71 AC027136 AC027136 Staphyloc
15 227 24.0 2815 2 AF269779 AF269779 Staphyloc
16 226 23.9 147006 84 SPNEU1904 AF44926 Streptoco
17 225.8 23.9 2111 9 A64844 A64844 sequence 1
18 192.4 20.4 44449 76 AC027830 AC027830 Staphyloc
19 192.4 20.4 49487 76 AC0274320 AC0274320 Staphyloc
20 192.4 20.4 55389 1 AC025592 AC025592 Staphyloc
21 189.4 20.4 55389 1 AC027893 AC027893 Staphyloc
22 189.4 20.4 55389 1 AC027893 AC027893 Staphyloc
23 189.4 20.4 55389 1 AC027893 AC027893 Staphyloc
24 138.4 14.9 2453 2 AF270310 AF270310 Staphyloc
25 59.2 6.3 3883 2 LBA5118 AB016431 Staphyloc
26 58.2 6.2 619 1 AF015751 AF015751 Lactococc
27 57 6.0 213732 1 AE001862 AE001862 Thermococ
28 56.4 6.0 21343 1 AE001744 AE001744 Thermococ
29 53.8 5.7 6951 1 AF001955 AF001955 Streptoco
30 50 5.3 7218 10 AC004157 AC004157 Sequence 14
31 43.4 4.6 169546 60 AC004157 AC004157 Plasmodi
32 39.4 4.2 94239 13 AF014P22 AL137082 Arabidops
33 38.6 4.1 87835 12 AB005237 AB005237 Arabidops
34 38.2 4.0 161528 74 AC006436 AC006436 Homo sap
35 38.2 4.0 172064 90 AL445195 AL445195 Homo sap
36 38.2 4.0 175012 74 AC069072 AC069072 Homo sap
37 38.2 4.0 193378 64 AC015991 AC015991 Homo sap
38 38.2 4.0 193858 65 AC016760 AC016760 Homo sap
39 38 4.0 2825 6 THESPAGI M63017 Theileria a
40 38 4.0 168778 92 HSA490024 AL11338 Human DNA
41 38 4.0 189256 81 AL391004 AL391004 Homo sap
42 37.8 4.0 140064 13 ATAC009177 AC009177 Arabidops
43 37.6 4.0 195721 71 AC027059 AC027059 Homo sap
44 37.4 4.0 85356 13 ATAC49 AL080318 Arabidops
45 37.4 4.0 99814 13 AF0812 AL132969 Arabidops

ALIGNMENTS

RESULT 1
LOCUS AX006200 945 bp DNA PAT 24-NOV-2000
DEFINITION Sequence 1 from Patent WO9904016.
ACCESSION AX006200
VERSION AX006200.1 GI:9929071
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis.
REFERENCE 1. (bases 1 to 945)
Bacillus/Staphylococcus group; Bacillus.
AUTHORS Estell, D.A.
JOURNAL Patent: WO 9904016-A, 1 28-JAN-1999;
GENENOR INTERNATIONAL P.V (NL); ESTELL DAVID A (US)
FEATURES
Source 1. 945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 288 a 180 c 232 g 245 t
ORIGIN

Query Match 100.0%; Score 945; DB %; Length 945;
Best Local Similarity 100.0%; Pred. No. 1.2e-250;
Matches 945; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 atgacgaactgaacgctatttttcgaagcgtgtttcaagaagaatttggggcggacc 60
b 1 ATGACGACTGAAACGCTATTITTCAGCCGTTTTCAGAAAGAAATTGGGGCGGACC 60
Y 61 gctttagctgatttggctataccattccgtcaacaagacaggaggagtgtgggctttt 120
|||||

Db 61 GCTTTAGCTGATTTTGGCTATACCATTCCTCCGCACACGAGGGAGTCTGGCTTTT 120
Qy 121 gccgcgcatcaaaatgctcaaaagcgtgtgttcaaaaacggaatgataaagggttcacgctc 180
Db 121 GCCGCGCATCAAAATGCTCAAAAGCGTGTGTTCAAAACGGAATGTATTAAGGGGTACGCTC 180
Qy 181 agcgaattatgggaacatcacacacacattatctcggaacagcttgaaaggagacgtttccct 240
Db 181 AGCGAATTATGGGAACATCACACACATTATTCGACACACTTGAAGGGGACCGTTCCT 240
Qy 241 ctgcttaacaaatattagatgtctgaacgaactatctgttcgaatgacatccgaatgat 300
Db 241 CTGCTTACCAAAATATTAGATGCTGACGAGACTTATCTGTTCAGTCACTACATCCGAAAT 300
Qy 301 gaatalgcccacatacatalgaanaacggttgagcttggaaaaacagaatgcttggatcatt 360
Db 301 GAATATGCCAACATACATGATAAACCGGTGAGCTTGCAAAAACGAATGCTGTGATATTAT 360
Qy 361 gattgccaacaaatgctcgagattatttatttgccacaatgacacaacaaagaagaacta 420
Db 361 GATTGCCAAAATATGCTCGAGATTATTATTCGCCACAAATGCAACAAAGGAAGAACTA 420
Qy 421 actacacatgataagcgttgagatgagatgagatgagatgagatgagatgagatgagatg 480
Db 421 ACTACCATGATGAGCGTGAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 480
Qy 481 ggggatttttctctatgctgcaagcggatcttgcattgacgattgaaaggaattcttgc 540
Db 481 GGGGATTTTCTATGTCGACACCGGTACTGTTCATGCAATGGAAGAAAGAAATCTTCT 540
Qy 541 ttgagacgcgacagacactcaagacacacactcaatattatgatatgacgcgaagaat 600
Db 541 TTGAGACGCGACGACGACACCTCAGACACACCTCAGATTTATGATTTATGACGGAAGAT 600
Qy 601 ggcgaagcgaagcgtcgcgagcgtctcatctgaaaaaagagatgaaatgaaatgaaatg 660
Db 601 GCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 660
Qy 720 tctatccgaagaagcgtatcagttcaccatgaaacaaatgaaatgaaatgaaatgaaatg 720
Db 720 TCTATTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 720
Qy 721 tctatgaaatgctcactcttgcgtgaggaatgaaatgaaatgaaatgaaatgaaatg 780
Db 721 TTGATGTAATGCGCTTACTTTCGTTGAGGGAATGCAACTTATGAGATGACGACGACG 780
Qy 781 aagcgaacaaacacatctctctatcagtgatgaaagggagggcgtatgactct 840
Db 781 AAGCAGCAAAACCATCTCTCTATCAGTGTGATGAGGAGGCGCGATGATCTCT 840
Qy 841 ggtgagatgctctatctcttcaaaaagagatcatatgtctgctccttaaggtctga 900
Db 841 GGTGAGATGCTCTATCTCTTTCAAAAAGAGATCATATGTTGCTTACGCTTTCGA 900
Qy 901 gaattaaactgaaagatatgagaatgtatcgtctcccatctg 945
Db 901 GAATTAACTCGAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 945

RESULT 2
LOCUS BSUB0007 216750 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 7 of 21): from 1194391
ACCESSION 299110 AL009126
VERSION 299110.1 GI:2633472
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis.
REFERENCE 1. (bases 1 to 216750)
Bacillus/Staphylococcus group; Bacillus.
AUTHORS Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,

gene	acetyltransferase	note="alternate gene name: argA, argB; ornithine acetyltransferase"
CDS	1176..2396	/transl_table=11 /product="amino-acid acetyltransferase" /protein_id="CAB12977.1" /db_xref="GI:2633474" /db_xref="SWISS-PROT:P36843" /translation="MIOSEDIKRVYTDVDSKPFQAKYCHGICRYSKDGIYITSE TRVSAVAYTOSHFAPIKATYQDOLKHEGPIKALYVNSALNCTGBOGICADYTR ESFASOLGIEPELVVSSSTGYVIGLHDEKIHAGIELKIEPSSGDESEKILITDLY IKOTYELALIGKTYITVIGARGSGMIPNNATMGFPTDAAIEEALQALRETDY VSFNQITVAGETSDMWLVYVMAACAENECECTEDHPDPVKALLICEDLAEIAR DEGAGTKLIEAOVCAKNNMLNVAIVAKIVGNSLVKAVVGTDNWRIIGAIHSAQA QYTAEEVEVYLQDCLFRNNPEPSESIAKEYLEGDEITIVIKMAGDNGRAMGCD LYVDYIKIASYRT" 2411..3187 /gene="argB" /EC_number="2.7.2.8" /function="arginine biosynthesis" /codon_start=1 /transl_table=11 /product="N-acetylglutamate 5-phosphotransferase" /protein_id="CAB12978.1" /db_xref="GI:2633475" /db_xref="SWISS-PROT:P36840" /translation="MKRTIVFKGGSVYIRELSEEFYONLKELRASGKWLIVHGSGPE ITNMKLRIKITEFGSGGKRTKPYALEVAEWLVSGCVKPFVAELKRAALIAISK DGGLLEADYLEDPEYGEVGEIKRVKDVASVNALEMNGIIPVIAPILSMDSCKTLVNDML LASAVAGALEADKLMFVETVDVGIKKEQRLDVLTKEIQMLINQEVITTGMIKPKVNS ALSALDSQVSEVIVNGKSPFAEQTFQGTIKVKAQVNS" 3184..4341 /gene="argD" /EC_number="3.184..4341" /gene="argD" /function="2.6.1.11" /function="arginine biosynthesis" /codon_start=1 /transl_table=11 /product="N-acetylornithine aminotransferase" /protein_id="CAB12979.1" /db_xref="GI:2633476" /db_xref="SWISS-PROT:P36839" /translation="MSLSFQTYGRMDIDIKRAKGTVEVDONKTYLIDIGIAVSNLIG HCEAVTEAVKROLDSEVHVSNSLQEQASADILVFPNCSQNGEKG KYLAKRATGKTKITPLFQSGHRTYAGMATGQDIKTFGPMLGFEYLPNDPSAE KALGEGDIAAVALTEVQSGGVNPAABEFLSAVOSCEKQOALLIDEIQTGIRGQ KGFAYHGPISPIITVAKGLNGSPVGVATSKQLEAFYTGSGTTPFGNNMLAAMAP VNTLQIVPOPEFLGKAADKGAFLKEQLEAEKLSFVKQIRGKGLMIGIEDGVPADY IAELQTLGILYIPAGSERDSAAPILYITNDEIAEAVSKQAIAHSAVNS" 4412..5473 /gene="carA" /EC_number="4.12..5473" /gene="carA" /function="6.3.5.5" /function="arginine biosynthesis" /note="alternate gene name: cpa" /codon_start=1 /transl_table=11 /product="cataboyl-1-phosphate transferase-arginine (subunit A)" /protein_id="CAB12980.1" /db_xref="GI:2633477" /db_xref="SWISS-PROT:P36838" /translation="MEGYLVLEDGTSFSGELDGHENGCEAVFETGATGQYEVLLDPS YKQIIVETYPILGNGINDEKDESKPQVKAAYVACDHRSHSEAYVSLKEYLOAKM NPLILHVDHRAVKKIRANGTGKATVVPQOMEAIVAIKPDGIVLNSNGDPAIK NKHIALIDRGYKKSISLVKRGCKATVVPQOMEAIVAIKPDGIVLNSNGDPAIK PYLGIKTSISREPRTGICGHOILALAGSGVNFPIPGHRAHNPVDIRKTKREVM SQRHSYVDDQSIINEEILTRPHIVNDISYVBLAKKILPYMSVQYHPRAHGCAASESE IFDQYLANVYIARREIATA"
gene	denhydrogenase	note="alternate gene name: argA, argB; ornithine acetyltransferase"
CDS	1176..2396	/transl_table=11 /product="amino-acid acetyltransferase" /protein_id="CAB12977.1" /db_xref="GI:2633474" /db_xref="SWISS-PROT:P36843" /translation="MIOSEDIKRVYTDVDSKPFQAKYCHGICRYSKDGIYITSE TRVSAVAYTOSHFAPIKATYQDOLKHEGPIKALYVNSALNCTGBOGICADYTR ESFASOLGIEPELVVSSSTGYVIGLHDEKIHAGIELKIEPSSGDESEKILITDLY IKOTYELALIGKTYITVIGARGSGMIPNNATMGFPTDAAIEEALQALRETDY VSFNQITVAGETSDMWLVYVMAACAENECECTEDHPDPVKALLICEDLAEIAR DEGAGTKLIEAOVCAKNNMLNVAIVAKIVGNSLVKAVVGTDNWRIIGAIHSAQA QYTAEEVEVYLQDCLFRNNPEPSESIAKEYLEGDEITIVIKMAGDNGRAMGCD LYVDYIKIASYRT" 2411..3187 /gene="argB" /EC_number="2.7.2.8" /function="arginine biosynthesis" /codon_start=1 /transl_table=11 /product="N-acetylglutamate 5-phosphotransferase" /protein_id="CAB12978.1" /db_xref="GI:2633475" /db_xref="SWISS-PROT:P36840" /translation="MKRTIVFKGGSVYIRELSEEFYONLKELRASGKWLIVHGSGPE ITNMKLRIKITEFGSGGKRTKPYALEVAEWLVSGCVKPFVAELKRAALIAISK DGGLLEADYLEDPEYGEVGEIKRVKDVASVNALEMNGIIPVIAPILSMDSCKTLVNDML LASAVAGALEADKLMFVETVDVGIKKEQRLDVLTKEIQMLINQEVITTGMIKPKVNS ALSALDSQVSEVIVNGKSPFAEQTFQGTIKVKAQVNS" 3184..4341 /gene="argD" /EC_number="3.184..4341" /gene="argD" /function="2.6.1.11" /function="arginine biosynthesis" /codon_start=1 /transl_table=11 /product="N-acetylornithine aminotransferase" /protein_id="CAB12979.1" /db_xref="GI:2633476" /db_xref="SWISS-PROT:P36839" /translation="MSLSFQTYGRMDIDIKRAKGTVEVDONKTYLIDIGIAVSNLIG HCEAVTEAVKROLDSEVHVSNSLQEQASADILVFPNCSQNGEKG KYLAKRATGKTKITPLFQSGHRTYAGMATGQDIKTFGPMLGFEYLPNDPSAE KALGEGDIAAVALTEVQSGGVNPAABEFLSAVOSCEKQOALLIDEIQTGIRGQ KGFAYHGPISPIITVAKGLNGSPVGVATSKQLEAFYTGSGTTPFGNNMLAAMAP VNTLQIVPOPEFLGKAADKGAFLKEQLEAEKLSFVKQIRGKGLMIGIEDGVPADY IAELQTLGILYIPAGSERDSAAPILYITNDEIAEAVSKQAIAHSAVNS" 4412..5473 /gene="carA" /EC_number="4.12..5473" /gene="carA" /function="6.3.5.5" /function="arginine biosynthesis" /note="alternate gene name: cpa" /codon_start=1 /transl_table=11 /product="cataboyl-1-phosphate transferase-arginine (subunit A)" /protein_id="CAB12980.1" /db_xref="GI:2633477" /db_xref="SWISS-PROT:P36838" /translation="MEGYLVLEDGTSFSGELDGHENGCEAVFETGATGQYEVLLDPS YKQIIVETYPILGNGINDEKDESKPQVKAAYVACDHRSHSEAYVSLKEYLOAKM NPLILHVDHRAVKKIRANGTGKATVVPQOMEAIVAIKPDGIVLNSNGDPAIK NKHIALIDRGYKKSISLVKRGCKATVVPQOMEAIVAIKPDGIVLNSNGDPAIK PYLGIKTSISREPRTGICGHOILALAGSGVNFPIPGHRAHNPVDIRKTKREVM SQRHSYVDDQSIINEEILTRPHIVNDISYVBLAKKILPYMSVQYHPRAHGCAASESE IFDQYLANVYIARREIATA"
gene	denhydrogenase	note="alternate gene name: argA, argB; ornithine acetyltransferase"
CDS	1176..2396	/transl_table=11 /product="amino-acid acetyltransferase" /protein_id="CAB12977.1" /db_xref="GI:2633474" /db_xref="SWISS-PROT:P36843" /translation="MIOSEDIKRVYTDVDSKPFQAKYCHGICRYSKDGIYITSE TRVSAVAYTOSHFAPIKATYQDOLKHEGPIKALYVNSALNCTGBOGICADYTR ESFASOLGIEPELVVSSSTGYVIGLHDEKIHAGIELKIEPSSGDESEKILITDLY IKOTYELALIGKTYITVIGARGSGMIPNNATMGFPTDAAIEEALQALRETDY VSFNQITVAGETSDMWLVYVMAACAENECECTEDHPDPVKALLICEDLAEIAR DEGAGTKLIEAOVCAKNNMLNVAIVAKIVGNSLVKAVVGTDNWRIIGAIHSAQA QYTAEEVEVYLQDCLFRNNPEPSESIAKEYLEGDEITIVIKMAGDNGRAMGCD LYVDYIKIASYRT" 2411..3187 /gene="argB" /EC_number="2.7.2.8" /function="arginine biosynthesis" /codon_start=1 /transl_table=11 /product="N-acetylglutamate 5-phosphotransferase" /protein_id="CAB12978.1" /db_xref="GI:2633475" /db_xref="SWISS-PROT:P36840" /translation="MKRTIVFKGGSVYIRELSEEFYONLKELRASGKWLIVHGSGPE ITNMKLRIKITEFGSGGKRTKPYALEVAEWLVSGCVKPFVAELKRAALIAISK DGGLLEADYLEDPEYGEVGEIKRVKDVASVNALEMNGIIPVIAPILSMDSCKTLVNDML LASAVAGALEADKLMFVETVDVGIKKEQRLDVLTKEIQMLINQEVITTGMIKPKVNS ALSALDSQVSEVIVNGKSPFAEQTFQGTIKVKAQVNS" 3184..4341 /gene="argD" /EC_number="3.184..4341" /gene="argD" /function="2.6.1.11" /function="arginine biosynthesis" /codon_start=1 /transl_table=11 /product="N-acetylornithine aminotransferase" /protein_id="CAB12979.1" /db_xref="GI:2633476" /db_xref="SWISS-PROT:P36839" /translation="MSLSFQTYGRMDIDIKRAKGTVEVDONKTYLIDIGIAVSNLIG HCEAVTEAVKROLDSEVHVSNSLQEQASADILVFPNCSQNGEKG KYLAKRATGKTKITPLFQSGHRTYAGMATGQDIKTFGPMLGFEYLPNDPSAE KALGEGDIAAVALTEVQSGGVNPAABEFLSAVOSCEKQOALLIDEIQTGIRGQ KGFAYHGPISPIITVAKGLNGSPVGVATSKQLEAFYTGSGTTPFGNNMLAAMAP VNTLQIVPOPEFLGKAADKGAFLKEQLEAEKLSFVKQIRGKGLMIGIEDGVPADY IAELQTLGILYIPAGSERDSAAPILYITNDEIAEAVSKQAIAHSAVNS" 4412..5473 /gene="carA" /EC_number="4.12..5473" /gene="carA" /function="6.3.5.5" /function="arginine biosynthesis" /note="alternate gene name: cpa" /codon_start=1 /transl_table=11 /product="cataboyl-1-phosphate transferase-arginine (subunit A)" /protein_id="CAB12980.1" /db_xref="GI:2633477" /db_xref="SWISS-PROT:P36838"

Query Match	100.0%	Score 945	DB 2	Length 216750
Best Local Similarity	100.0%	Pred. No. 26-250		
Matches 945	Conservative 0	Mismatches 0	Indels 0	Gaps 0
1 atgaagactgaacgcttattttcaagccgcttttcaagaagaagaaattggagcgagacc 60				
ATGAGAGACTGAACCGCTTATTTTTCACGCGCTTTCACAAAGAAATTTGGCGCGGAC 79833				
61 gcttagctgaatttgggtcctacaccctcgctcgaagaagaagggagtgcttggtctt 120				
GCTTAGCTGAAATTTGGCTATACCATTCGCGACAAAGCAAGGAGAGTGCAGGCTTT 79833				
121 gccgcgacatcaaaatgggtcgaagcggttgctcaaaagaagatgctataaggggttcagctc 180				
GCGCGGACATCAAAATGGTCAAAAGCGTGTTCACAAAGCGAATGTATTAAGGGGTTCACAGCTC 79953				
181 agcgatatatgggaacatcacagaacatttttcggagcaagcttgaagaaggagaccgtttccct 240				
AGCGATATATGGGAACATCACAGACATTATTTCGGACAGCTTTGAAGGGACCGCTTTCCCT 80013				
241 ctgcttcaaaaaattatgatagtctgaccaggaacttatcgcttaagtgcacgcgaatgat 300				
CTGCTTCAAAAAATTATGATAGTCTGACCGAGGACTTATCTGTTAAGGTGCATCCGAATGAT 80073				
301 gaatatgccaacatcatatgaataacggtgagctctggaaaaacagaatgctgtgtaacattt 360				
GAAATATGCCAACATCATATGAATAACGGTGAAGCTCTGAAAAACAATAAGCTGTGTAATATT 80133				
361 gattgccaaaaagatgccgagatattattlaagggccacatgccaacaaagaaggaagaca 420				
gene	5466	.8549		
gene"carb"	5466	.8549		
CDs	5466	.8549		
gene"carb"	5466	.8549		
EC.number="6.3.5.5"	5466	.8549		
function="arginine biosynthesis"	5466	.8549		
note="alternate gene name: cpa"	5466	.8549		
codon_start=1	5466	.8549		
transl_table=11	5466	.8549		
product="carbamoyl-phosphate transferase-arginine	5466	.8549		
(subunit b)	5466	.8549		
protein_id="CAB12981.1"	5466	.8549		
gb_xref="GI:263478"	5466	.8549		
db_xref="SWIS-Prot:P18195"	5466	.8549		
translation="MPKDSISSILVIGSPITIGQAEEFDYSGTGGCIATKEGCRYR	5466	.8549		
EVNSNPATIMDEAFADETIYPEPILPAESLTATIKERPDGILANLGGOTYLNLAIVE	5466	.8549		
ETVNLHEGVKLLGTISVETIYKGEDREKESRLIMELKQPVESNIVENDALAHVE	5466	.8549		
SIGPVIRIPAYVLTKGAGGAGIADKFAFPAKMLASPIVOCLEVESISIGFEIE	5466	.8549		
EVMSDNNVITICVCMENIDPVGVHGDSDIVAAPEVHSDKATGYPAKMAALAVGYLDEI	5466	.8549		
VNGCSTOFALDFPSKOYIVLEVPVPSKSNADKLTGMATGEVNALEMYLDEI	5466	.8549		
KNPITGSTVAFPSKOYIVLEVPVPSKSNADKLTGMATGEVNALEMYLDEI	5466	.8549		
IKQKGALELKNIGTHLPBLSGLSIDTMDLAIPTPDRFEVYVMEILSRYSIDIE	5466	.8549		
KTADPFLETPFNIIKLEKRLMEAGSDSPILKKAREGSDATIASLSTKEE	5466	.8549		
RLRKEGDFIPSEKIVDTCAEPDAKTANFYSTYEGDGDGSRKKRALLISGP	5466	.8549		
RIGCGVEFDYSAHVGLTLOELFEETIMNNPVEYSDYETIADLYPEPMTTHIYN	5466	.8549		
VREENIDFALVOEGGOTANMAEALERKGTLLGSPOTDVEDRPOFLDEIG	5466	.8549		
LAHKGIEATIKREAKSAKEIGYPALIPSVIGMGKLYVDSQALSOILNDEDSG	5466	.8549		
PYGLIQQVYSGKEVIDISLQISNDEYFNTYEHLIENVRASRYTPVAKSGWPRAND	5466	.8549		
QOQPKDIAOKIARKLSFKGIQIMQEVFIONGLTILEVNRASRYTPVAKSGWPRAND	5466	.8549		
RLAGASLKLNDLPANOVNHHGVAKPFVPSHMAIOPYDXTLGPENKSTGEGMCVADSN	5466	.8549		
SALKIATRYMSOKGSIYLDNVEYKELIENAGDTIHGTPASMDGEGSLHINLS	5466	.8549		
GSEPAKREKEAMTHGIPVTEETVRAITLQSGSHPPQVSLDKLKEVASCQ"	5466	.8549		
8537	9510			
gene"argp"	8537	9486		
CDs	8537	9486		
gene"argp"	8537	9486		
EC.number="2.1.3.3"	8537	9486		
function="arginine biosynthesis"	8537	9486		
/codon_start=1	8537	9486		
/transl_table=11	8537	9486		
/product="ornithine carbamoyltransferase"	8537	9486		
/protein_id="CAB12982.1"	8537	9486		

Db	80134	GAAGCCCAAAAAGATGCCGAGATTATTATTGGCCACATTGCAACAAACAAAGGAAGACTA	80193
Qy	421	actccaccatgtagagcgctggaagaatgtagatgtagctctggccgtgtgaaaggtgaaacgg	480
Db	80194	ACTACCATGATTAAGCGCTGAGCAATGAGATGAGCTCTTGGCCGCTGTAAAGCTAAAGCCG	80253
Qy	481	gggggatttttctatctgtgccaaagcgttactgttctacgatttgaaaaagaaatctctgct	540
Db	80254	GGGGATTCTTTCTATGTGGCCAAAGCGGTACTGTCTTCAATGGCGATTGGAAAGGAATCTTGCT	80313
Qy	541	ttggaagcgcagcagaactcgaacacaactacagattatgatgatgatgatgacccaagaat	600
Db	80314	TTGGAGACGCGACGACGAACCTCAGACACAACCTCAAGATTATGATGATTATGACCCGAAAAGAT	80373
Qy	601	gcgaagaagcgaagctgcgcgagcttcatctcgaaaaaagaagcattggaagtgataggtcccg	660
Db	80374	GCACAAAGCAAGCTGGCGCAGAGCTCATCTCTGAAAAAGACATTTGAAGTATGATGAGAGTCCGC	80433
Qy	661	tctcttcgaagaagcgatcacacgtcacgtcacctgacacaatttggagatttggcttaacagaca	720
Db	80434	TCTTTCTCCAAACAGCGATACAGTCCACGTCCATGAAACAAATTGAGATTGGCTTACACACACA	80493
Qy	721	ttgatctaacgcgccttactcttcgcggtcggggaatggaatgaaacttaccagatcagcaagtta	780
Db	80494	TTGATCTAAATGCGCCTTACTCTTTGGTGGGGAATGGAACCTTATGACGATCGCAAGCTTA	80553
Qy	781	aagcagcgaagaacacctctctctctcttcagttcagttgattggaaggggagcgctatattct	840
Db	80554	AAGCAGCAAAAACCATTTCTCTTATACAGTGTGAATGAAGGGAGGCGCCGATGATCTCT	80613
Qy	841	gtgtacgattctcattcattcttcttcttcttcttcttcttcttcttcttcttcttcttcttga	900
Db	80614	GGTGAAGTATGCTTATGCTTTTCACAAAAGCAAGATCATATGCTGCTGCTGCTGCTGCTGCTGA	80673
Qy	901	gaatttaactcgaagagatgacagaagatgatactgctctccatctg	945
Db	80674	GAAATTAACTCGACAGATATGACGAATGTATCGTCCCACTG	80718
RESULT	3		
LOCUS	D88802	36308 bp	DNA
DEFINITION	Bacillus subillus		BCT 07-FEB-1999
ACCESSION	D88802		
VERSION	D88802.1		GI:1945083
KEYWORDS	RNA-Asp; RNA-Met; RNA-Gly; RNA-Arg; 5S ribosomal RNA; 16S ribosomal RNA; 23S ribosomal RNA; yidJ; groEL; groES; yidL; yidK; yidJ; yidL; yidH; yidE; yidF; yidA; yidC; yidB; yidA; gly-TRNA; Arg-TRNA; ydhH; ydhT; ydhs; ydhr; ydhQ; ydhP; ydhO; ydhn; ydm; ydhL; ydhK; ydhJ; ydnt; ydnt; ydhG; ydhG; ydhF; ydhE; ydhD; ydhC; ydhB.		
SOURCE	Bacillus subillus (sub-species:Marburg, strain:168, isolate:JH642)		
ORGANISM	DNA.		
REFERENCE	Bacillus subillus		
AUTHORS	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.		
TITLE	1 (bases 1 to 36308)		
JOURNAL	Sadate,Y., Yata,K., Fujita,M., Sagai,H., Itaya,M., Kashihara,Y. and Ogawara,N.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (05-NOV-1996) to the DDBJ/EMBL/Genbank databases. Yoshitoku Sadate, National Institute of Genetics, Radiosotope Center, Yata 1111, Mishima, Shizuoka 411, Japan (E-mail:y.sadate@lab.nig.ac.jp; Tel:81-0553-81-6870, Fax:81-0553-81-6870)		
TITLE	2 (sites)		
JOURNAL	Sadate,Y., Yata,K., Fujita,M., Sagai,H., Itaya,M., Kashihara,Y. and Ogawara,N.		
REFERENCE	Nucleotide sequence and analysis of the phoB-rnE-groESL region of the Bacillus subillus chromosome		
AUTHORS	Microbiology 143 (Pt 6), 1861-1866 (1997)		
FEATURES	97346038		
source	Location/Qualifiers		
	1..36308		

gene
CDS
/organism="Bacillus subtilis"
/isolate="JH642"
/strain="168"
/sub-species="Marburg"
/db_xref="taxon:1423"
1..866
/gene="ydgk"
<1..866
/gene="ydgk"
/note="B. coli bicyclomycin resistance protein; P28246 (694) transmembrane"
/codon_start=3
/transl_table=11
/protein_id="BA19692.1"
/db_xref="GI:1945084"
/translation="AGLYTSRAIVRDVGTGREGTSKEPSSLMVTAVAPVAPMTGCAI
LLPAPVMTHTFHYLMITIGELVLLALRKETPLPLEKRIPSSIGTSYTKSLKDR
SFMGALVTGFIHGSRFVSGTFVYQDITGVSPVSLFGISGLAIISSFFIIG
FEGITHESRLAVITAMIAVTLMTMIHGPALVISTITWTIGMVLSTFT
LAMEGQRRASASALGMLPLLGISTVSPVGINETTVAPMGALMFAVIGSLAF
GLTERVQNS"
complement(861..1640)
/gene="ydhb"
/note="transmembrane"
/codon_start=1
/transl_table=11
/protein_id="BA19693.1"
/db_xref="GI:1945085"
/translation="MLTIIVMLGIIIGTIGAGAGCAVTAITLLEPIPIHAIAGTS
LAGAPFTLSGAYSHVREGNTOMKIGLVGSPAGVSFGAKTSTFTIPDIAHVLG
MLFSAILILRLPLEKRAVNSTISTYTRAVIGLISAGVLSFGIGSAPFTDIG
LMLMLNLSIRSVGTMVLVITPELVAGVGIGITGEGVDYLLVAVLGMTCAGVYAK
FTMLNPKVYVLSHLPDSGDCRAVVLTKTKMLTSCVPL"
1899..2563
/gene="ydhc"
1889..2563
/gene="ydhc"
/note="B. subtilis gluconate operon repressor; P10585 (209) DNA binding"
/codon_start=1
/transl_table=11
/protein_id="BA19694.1"
/db_xref="GI:1945086"
/translation="MDKFKIDKPRVLYQFYNOLKKMIFNGTFKPERINETQIAKSF
GYSRSPLEARILKXGILKADNRNGFSITSLAKVNDIYKRIPIRLDAVEIYD
PADREKLTILKOLEPPEKALINGTEPTETIIRLNOKPHELVPSNNRKLNLPEVN
DLIHCRILNTGCHRAETIIRHRRIFEEYKKNKKAQKHYLAHPNDCHELIKVL
DEGKEN"
2690..4009
/gene="ydhb"
2690..4009
/gene="ydhb"
/note="B. subtilis yaan hypothetical protein; P37534 (704)"
/codon_start=1
/transl_table=11
/protein_id="BA19695.1"
/db_xref="GI:1945087"
/translation="MFIHVGPDSLPSIGRCAGVADQIRGVNGLDENTIVGQAL
IPLYVYVOPDRLTALAIAKAFVPLERLAANPGISPNALOGAKITIPISVINGT
LSFYLRMPDDRELINDYAYSSSISIEFYHATPNCIDINOLDAALITPMORAVT
PLATYNTSGSFREIYHOVLNNSTARTVNNYVITDYSRGYCACTIDFEVNSAD
RDLFGPLROLDRLOAGCYLRTAVPAKTSNDIPMRGVDYGICGVVNNYNTIAYD
MHAASGEPVPAITEIRTIETPIAYPSKKITIGVPLGIDMILITIOGIVASIS
NONAKREARQAPVIOYSAKOSPFPRISDQGRTHVMEEDVPSMKQIVREYRL
QAIQMADLTALRAHFGCNFLSECKKHHOSLVFLLKSAAE"
4155..5300
/gene="ydhb"
4155..5300
/gene="ydhb"
/note="S. lividans glycosyl transferase; J50636 (311)"

gene
CDS
/codon_start=1
/transl_table=11
/protein_id="BA19696.1"
/db_xref="GI:1945088"
/translation="MPTVIGITKASDKGVYHNTSTKTKYKRLKLEAGATVYHLRDL
RTPIHVSFNGPGLDFVKTHITSTLIDIVADLSKIQFQPIYDKRGAELVRDL
DIPGVSSASFELEDFEHLKTLPLPESGAPLEIDECDDLAKMEYGVAPKMLVQF
MNNKGEILNVYVTSRYPSPESDRGDECLTIGSFPRKAEKDFPLEQIDKERVYISM
GVLIDHTEDEFNLICDAFSGFNGKVIYIAGEADLTAKQAPENFIAPVPLEVLE
QSDVEYIHGGSNNVNGIHFSFVPLVMPHDKQAPVORLSELNAGVYSDEVNAOII
LKQADVEILNDDYTGRIKINSQFECKMDEVARIDEILRKAK"
complement(5339..6103)
/gene="ydhb"
complement(5339..6103)
/gene="ydhb"
/note="membrane protein"
/codon_start=1
/transl_table=11
/protein_id="BA19697.1"
/db_xref="GI:1945089"
/translation="NSRCLEYTGSLFNGGIIIMRILSLVPAIMLACSSNASTEKRO
HAGEKTVKAPQSTSSQKSDTDVQPNQVTDRTLRKVGQTSDDKRAVLNDIKQ
VNPKYKIGDVELTVKEMKLIHLRPDYSMDIYHEHLEHDEDFPVKVPVDIYNSTKRV
NVAPIALMKTNGTEFDMNKDLIELNELNELGEGKEGNGLGFIVNASGHAHKAAD
AEKRTKEIKWIEITITSDVDFDHNKRIKSAQKIKIF"
complement(6115..7503)
/gene="phob"
complement(6115..7503)
/gene="phob"
/note="B. subtilis alkaline phosphatase IIA; P19405 secretory"
/codon_start=1
/transl_table=11
/protein_id="BA19698.1"
/db_xref="GI:1945090"
/translation="MKRFKRLPLPAVLSIAFSSLASGVPEASQERKKGNODEIKR
NVILIDIGGVSTYSAYRYLKDNRKTVVEPTADQYLVGQTTYPDDEQWNTYDASA
AAATAMSGAGIKITYNNAIANDNGSEAKVITLEAKRSGNGLVNAISELTHATPASFSS
HDSHRKNMNSIADYDFDEYVNGSKITDVLIGGKSNFQKBNLTKERKAGSYVDD
RKDMLKKNKDSQVGLADGLPKRIIDRTYDILSGLKNDNTALIKLNDKDFPLFMVNG
SOLIDMAGHNDIVGAMSEMDPEOAYKKAIDPAKDKHTLVVATDHSHTGYSIGASDS
IYNWSEPIKAAKRTPDFAEKIADGADVETIKTYIDOKIALTLAKIQAIEBAAKSF
KEVDIDNALIENIFENKRSHTGWTGTGEDVPVYAYGSSSETGAGIDNTEIAKNVF
KALQYNIKINDK"
7829..8236
/gene="ydhb"
7829..8236
/gene="ydhb"
/note="B. subtilis yaan hypothetical protein; P46328 (704)"
/codon_start=1
/transl_table=11
/protein_id="BA19699.1"
/db_xref="GI:1945091"
/translation="MEMHFTLGGNMADVSEYLAGIADPPHRETEYVLTWIKNKYE
NLHETIKWNPMTDGTFTIGSVSKRLAAVAPKVTIAHVEDDIVAGADYVTEQLL
RIPWNGVDYTLLEKMLEFNLKADCSFTWKK"
complement(8312..8809)
/gene="ydhb"
complement(8312..8809)
/gene="ydhb"
/note="secretory"
/codon_start=1
/transl_table=11
/protein_id="BA19700.1"
/db_xref="GI:1945092"
/translation="MSHRYKPLIPTAFILIAFCLIFESYKLIHVRLPEYNNQSESK
DOMEAVFTINEDAPNRSGLKWIQDITLIDITVLTLSLTKDVEYLSDDTEIPMA
DYAGGTFSDSAKESVSEFLERDHEHGHDTTEVKKQGNHISASQTLNENKTL
DEKQ"
complement(8827..9309)
/gene="ydhb"
complement(8827..9309)
/gene="ydhb"
/note="B. subtilis yaan hypothetical protein; P46328 (704)"

```

(512)"
/codon_start=1
/translation=1
/protein_id="BA19701.1"
/db_xref="GI:1945093"
/translation="MIIIPNNIEAKHMLTDFEAKHMGCTPEMAISSGIFRCDELGFA
AVNESEIGICITITITDCKDEIISDSRMTLEOPALCADRDYTLDMYRGH
TNDNNNAIAFQKKNYOPAAVFPNAVEKARLKEIFEIVANGILIDELLESKVID
"
gene
9392. 10369
/feature="ydnv"
9392. 10369
/feature="ydnv"
9392. 10369
/feature="ydnv"
/codon_start=1
/translation=1
/protein_id="BA19702.1"
/db_xref="GI:1945094"
/translation="MEISDIIRYGOHHIDGVLLEELIKSAPVORLKGIIYOGASPLVNRK
MNTVREHSIGVWLKIKLGTIEBOIGLHDSHTAFSHVDVDFENOAEDYHNT
FOOVIVHSIPILOKHGHTHEELISDSRMTLEOPALCADRDYTLDMYRGH
INLHEMOTRIDLIYNGRMPDSTEAEMFVSYYKKEVIDPFTNPVAVGXYLARA
LKAALRHVYISADLKTDEVLNLRKSKNEVSLSTSHPGIYIEDDIOYDFHO
Query Match 33.9% Score 320.4; DB 2: Length 36308;
Best Local Similarity 59.4% Pred. No. le-77;
Matches 562; Conservative 0; Mismatches 381; Indels 3; Gaps 1;
QY 1 atgagacgtgaacgcttatcttcaagcctgtttcaagaagaatttggggcgggagc 60
Db 17823 ATCATGACGCGATCATATTTTATGAGCGCTGCTTTAAAGAACTATGAGGAGGAGC 17882
QY 61 gcttaagctga---tttggctataccatccgtcccaagaacaggagtgctggc 117
Db 17883 AAGCTTGCGTGGCTTTTGGCTAGCGCAATACCTCACAACAAACAGGTAGTGGCTGGCC 17942
QY 118 ttggcggcgatcaaaatgctcaagcgtgtgtcaaaaggaatgataagggttcaag 177
Db 17943 GTTCTGACATGCCCCATGCTGCTGCTGTGTAAATAAGGCCCGCTGCAGAAAGACA 18002
QY 178 ctcaagcaattatgggaacatcaagaacattatctggaaagcttgaaggggagccgttc 237
Db 18003 CTTCATCAATATGGAAGATGATCATCCAGAGATATTCGGGTTCGCCGATGTAAAGGTGTT 18062
QY 238 cctctgcttaaaaaatatagatgtctgcaagaacattatctgttaagtgatccgaat 297
Db 18063 CCGCTGCTGTAAGGCTGTCGAGCGCAATATGATGATCTCCGTCGAAGTCCATCTGAT 18122
QY 298 gatgaatataccaaatacaatgaagaagtgagcttggaaacagaatgctgtacatt 357
Db 18123 GATGATATGCAAAACGTCGACGAAATGGGACCTTGTAAGGAGATGCTGTATATC 18182
QY 358 atgattgcacaaaagaatgcgaagatattatattgacacaaatgacacaaaagaaga 417
Db 18183 ATGATATGCAAAAGTGCAGCGCAACTAATTTTGGACATCATCAAGCAACAAAGGAAAG 18242
QY 418 ctaactacattatagacgttgaagaatggaagtccttgcgcgtgttaagtaag 477
Db 18243 TTCAAAACAAAGATAGAAACCGGTGATTTGGAACGGCTCTGAGCGAATTCAAAATCAAG 18302
QY 478 ccgggggagatttctctgtgcgaagcgtgactgctgcatgcatgttggaaagaatctt 537
Db 18303 CAGGAGATTTCTTTTATGTCACACGGTACACTCTTTATGTAAGGAAACCCCTT 18362
QY 538 gctttggagagcagcagaactcagacacacacattatagattatgacggaaga 597
Db 18363 GTCCCTGTAATTCAGCAAAACCTGTGATACAACTATGCGGTATACATTTATGACCGCTGT 18422
QY 598 gatgcagaagaagcagctgcgagctcatctgtaaaagaagattgaagtgatgagtc 657
Db 18423 AATGACCGAGCGCAAAACAACTCTTCATATGAGAAAGCAATGGAAGTCAATACAGATA 18482
QY 658 ccgtctatctcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 717

```

```

Db 18483 CCGCATATCGATAAAGTGCATACACCGAAGTAAAGAACTTGTAAACGCTGAGATCAT 18542
QY 718 acattgtgaatgcgcttacttctggttggaagaatggaactcagagatcagaagaagc 777
Db 18543 GTTATATGCAATCATGATATTTCTCTAGTGTACAAATGGAAGATGAGCGGAGAGTGT 18602
QY 778 ttaagaagcagaacacacatccctctctatcatcagtgatgagaaggagggcgctgctc 837
Db 18603 TTTCCTCATATCAAACTATTTGCTGGGAGTGTCTGAGCGGATCAGACAGATCATTA 18662
QY 838 tctgtgtagatgtctatcccttcaaaaagaagacatagatgttgcctcctaacgcttc 897
Db 18663 AATATGATATTCAGTATGAGATGCAATGCAAGCGCTCACACTTATTTGCTGGCAGATT 18722
QY 898 ggaagaattaaactcagaagatagcagaatgtatcgtccctccatc 943
Db 18723 GGAGAAATTTCAATAGAAAGACATGTAATTCATGATATCTCATC 18768
RESULT 4
LOCUS BSUB0004
DEFINITION Bacillus subtilis complete genome (section 4 of 21): from 600701 to 813890.
VERSION 299107 AL009126
KEYWORDS 299107.1 GI:2632866
SOURCE
ORGANISM
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Bacillus.
1 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertorello, M.G., Bessieres, P., Bolotin, A., Borczyk, S.,
Boris, R., Brouillet, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capano, V.,
Carter, N.M., Choi, S.K., Codan, J.J., Comerford, I.F., Cummings, N.J.,
Daniel, R.A., Denli, F., Devine, K.M., Dusterhoft, A., Ertlich, S.D.,
Emmerson, P.T., Entian, K.D., Erington, J., Fabret, C., Ferrari, E.,
Fougeron, D., Fritz, C.Y., Fujita, M., Fujita, Y., Funa, S., Galliz, A.,
Gallier, G., Gieseler, G., Guy, B.J., Hag, K., Haisch, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holst, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kashara, Y.,
Klaerner-Bianchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kunita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Mauch, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Nodack, N., Noone, D., O'Reilly, M.,
Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Patro, V., Pohl, T.M.,
Portetlelle, D., Portwoll, K.S., Prescott, A.M., Prescan, E., Fujic, P.,
Purcell, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Roche, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serron, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Yamakoshi, A., Tanaka, T., Terpe, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, P., Vassarotti, A.,
Vieri, A., Wambutt, R., Wedler, E., Wedler, H., Wetzinger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamanaka, K., Yasumoto, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
2 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45

```

68 89 48

FEATURES
source Location/Qualifiers
I . 213190
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
33 . 740
 /gene="ydfS"
33 . 740
 /gene="ydfS"
CDS
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12373.1"
 /db_xref="GI:2632867"
 /db_xref="SPTRMBL:P96697"
 /translation="MIELEVIKRTVASFGDLLIAERILGKQITISOMTIPEDFIAATLG
AIAAGRAVNTSIKPHNNAISIFVLTIPLISLKIKRKLRFPAGDPVLIANGKI
LESNRKMYRTYLLINOOLKEKEIFNEEFLFALETNGQLTVLRKPQGFHVKODIM
IVANOEGRPIELIDGEIIENNKLKNRRLTESMLLEEIRKRDKIKETVAVALGGND
IVDOYKRDHTSYMDKE"
complement(755..1186)
 /gene="ydtI"
complement(755..1186)
 /gene="ydtI"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12374.1"
 /db_xref="GI:2632868"
 /db_xref="SPTRMBL:P96698"
 /translation="MDFEKIRKKLEITNEYKOSDFTWNTLYKKAPEHPDEASTFY
DYRQDPENFIYEMGVREERTILSLSTOLLGTTPVPAPAEVEVIRENYIGCI
ERIQCPAEATHHLQTLGLNLGHLSYPQEVTAFNKGL"
complement(1199..1441)
 /gene="ydgA"
complement(1199..1441)
 /gene="ydgA"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12375.1"
 /db_xref="GI:2632869"
 /db_xref="SPTRMBL:P96699"
 /translation="MPQIINAIKINGVQNGINDVPYVHNSHTANSKYFCANFSL
GDSPISLTNGTSIHDSVDSDQDTGNPSAPISMQR"
complement(1453..1727)
 /gene="ydgB"
complement(1455..1727)
 /gene="ydgB"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12376.1"
 /db_xref="GI:2632870"
 /db_xref="SPTRMBL:P96700"
 /translation="WPSTVINLYLKINSISGNCSTIGEAAVNSPTNNOKSOCTNSS
FGDISPESVSEWFLNDPDVNDTSGNSDPTSVMNPAPIAPILD"
2026..2613
 /gene="ydgC"
2026..2613
 /gene="ydgC"
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12377.1"
 /db_xref="GI:2632871"
 /db_xref="SPTRMBL:P96701"
 /translation="MKSQGVSKRRRLTKAANEVSVGFDQAVSELTVKAGGTQPSF"

[illegible]

A F V G I E L V G T A G E T E N P E K V I P K A I N N I P V R L L F Y I G A L L V I M S T Y P W D I I N P S E S E
P F V Q V F V A V G I V G A S I I N F V L T S A S A C N S A V E F S T R M V Y S L A K D H N A P E S M A K L T T

Query Match	33.98;	Score 320.4;	DB 2;	Length 213190;
Best Local Similarity	59.48;	Pred. NO. 1.2e-77;		
Matches 562;	Conservative 0;	Mismatches 381;	Indels 3;	Gaps 1;

Query Match	33.9%	Score 320.4	DB 2	Length 213190
-------------	-------	-------------	------	---------------

Best local similarity 59.4%; Pred. NO. 1.2e-77;
Matches 562; Conservative 0; Mismatches 381; Indels 3; Gaps 1;

Best local similarity 59.4%; Pred. NO. 1.2e-77;
Matches 562; Conservative 0; Mismatches 381; Indels 3; Gaps 1;

QY	1	atgaacatgcaacccgttatttttcaagccgtgtttccaagaagaagattcttg99g9ggacc	60
Db	30821	ATCATGACGCAATCCATATTTTATAGAGCTGTCTTAAAGAAAGCATATGGGAGGAGACG	30880
QY	61	gctttaagttga---tttggctatacccttccgttaacaagaacaggggagatgctcggact	117
Db	30881	AAGCTTCGTGAGCGCTTTTGTGGCTACGCATATACCTCCATCAAAAACAGGTAAGTGGCTGGGCG	30940
QY	118	tttgcgcgcacatcaaatlgtctaaagcgtlgttccaacacggaatgtataaagggttcacg	177
Db	30941	GTTTCTGCACATATGCCATAGCGTCGTCTGTATAAAAATGAGCGCCGCTGGCAGAAAGACA	31000
QY	178	ctcaagcgaattatggaacatccacagaacattatctcgaagcagcgtgtgaaggggagccgttcc	237
Db	31001	CTTGATTCAAAGTATGAAAGATCATATCCAAAGATATCCGGGTTTCCGGATGCTATAGCTGTTT	31060
QY	238	cctctgcttacaataatttagatgctcaccagaactatctgtccaggttcacgtcgaat	297
Db	31061	CCGGCTGCTGTAACCTCTCTGGACCCCAATATGCAATCTCTCCGTCAAGTCCATCTCAT	31120
QY	298	gatgaatatgcacacatcacatgataaagctgtagctcttggaaaaacagaatgtctgtacatt	357
Db	31121	GATGATTTGCAAAACTCCACGAAATATGCGCCGCTCTGTATAAACCGAGTGCCTGTATATC	31180
QY	358	attgatctccaaaaaagacgcgcgagatttattatgtcccaatgtccacacaaagaagaagaa	417
Db	31181	ATTGATTTCAAAAGATGAGCGCGAATATATTTTGGACATCATATGCACAAAGCAAGAAAG	31240
QY	418	ctaactaccatgatagagcctgtggaagaatggatgagctctgtgcgcgcgttaagtaag	477
Db	31241	TTCAACAACAAGATATGAAAGCGGTGATTTGAAAGCGGCTCTCTAGCCGAATCAAAATCAAG	31300
QY	478	ccggggggatlltttctatgtgcgaagcggatctgttcatcgtattgaaagaagaattctt	537
Db	31301	CCAGAGAGATTTCTTTATATGTCACAAGCGGTACACCTCCATCGCTTATGTAAAGGAACCTT	31360
QY	538	gcttttggagacgcagcgaactcagacacaaactacagaattataltgatctatgacccgaaaa	597
Db	31361	GTCCTTGAAATCCAGCAAACTCTGTATCAACATATCCGATATTCAGTATATGACCGCTGT	31420
QY	598	gattcagaagaagcgaagctgtgcgagacttcatctgaaaaagagcatgaaagtatagatgc	657
Db	31421	AATGACCAAGGCGCAAAAAGAACTCTTCAATATAGAAAAAGCATGGAAGTCAATACGATA	31480
QY	658	ccgctctatctcagaagcggcatcacagttcacatgacaatctgaagattgtcttaacaag	717
Db	31481	CCGAGATATTCGATTAAGTGCATACACCGGAAGTAAAGAAAGTTGGTAACGCTGAGATCATTT	31540
QY	718	acatttatttgaatgcgtcttacttcttcggttgggaaatgaaacttatacgaagtcagcaagc	777
Db	31541	GTTTATGTGCAATAGATATATTTCTCACTGTATCAAAATGGAAGATATACGGCCGAGCTCT	31600
QY	778	ttaaagcagcaaaaaccctccctctctctcgtatgtatgaaaggaggagcgcatgtatgc	837
Db	31601	TTTCCCTTCATATCAAAACTATTTGTGTGGGAGGTGTTCGACGGATTCAGGACGATCATTA	31660
QY	838	ctctgttgatgtatgtcatccttcaaaaaaagagatatatgtgtcttcccttaagctt	897
Db	31661	AATATATGATATTCAGTATGATGATGCATATCCAGGCTCACACTTATATTCGCGCGCATTTT	31720
QY	898	ggagaattttaaactcgaagatatgcagaatgtatcgtctccatc	943
Db	31721	GGAGAAATTTTACATAGAAAGCAATGTGTAATTCATATGATATCTCAT	31766

	PAT	
AX006205	945 bp	DNA
LOCUS		
DEFINITION	Sequence 6 from Patent WO904016.	
ACCESSION	AX006205	
VERSION	AX006205.1	GI:9929073

DEFINITION Sequence 6 from Patent WO9904016

VERSION AX006205.1 GI:9929073

SOURCE

ACKNOWLEDGMENTS

REFERENCE

AUTHORS
TITLE

JOURNAL

FEATURES

① ② ③ ④ ⑤

BASE COUNT

ORIGIN

Query Mat

Query Match	33.8%;	Score 319;	DB 9;	Length 945;
Best Local Similarity	59.4%;	Pred. No. 1.7e-77;		
Matches 560;	Conservative 0;	Mismatches 380;	Indels 3;	Gaps 1,

QY 4 acgactgaacgctattttccaagcctgtttccaagaagaatttgggcyggaccgct 63

Db 1 ATGACGCATCCATTATTTTAGAGCCCTGCTTTAAGAGAAGACTATGGGAGGACGAAG 60

61	CTTCGTGAGCGCTTTTGGCTACGCCAATTACCCCTACAAAAAAGGTGAGTCTGGGCCGTT	120	
QY	121	gccgcgacatcaaaatggtctaaagcgttctcaaaacggaatgataaggggttcacgctc	180
Db	121	TCGCAACATGCCCAATGGCTGCTGCTGTATAAAAAATGGCCCGCTGGCAGGAAGACACTT	180
QY	181	agcgaattctgggaacatcacagacattattctggagacgcttgaaggggacgctttccct	240
Db	181	GATCAAGTATGGAAGATCATCAAGATATTCGGGTTCCGGATGGTAAAGGTTTCGG	240
QY	241	ctctctacaaaattatagatctgcaccaggaactattctctcagtgagatcagaattt	300
Db	241	CTGCTGTAAAGTGTGTGAGCCCAATATGAGATCTCTCCGCAAGTCATCTGTATGAT	300
QY	301	gaatatgccacatcatabgaaaacggtgagcttgtaaaaaacagaatgcctgtcatatt	360
Db	301	GATTATGCAAAATCTGACGAATAATGGCACTTGATTAACGGAGTCTGTATATCATT	360
QY	361	gattgccaaaagaatgpcgagatttctatgagccaatgcaacaacaaagaagaacta	420
Db	361	GATTGCCAAAGTAGCGCCGACACTATTTTGGGACATCATGCAACACAAAGGAAGATTC	420
QY	421	actacacatgatagagcgtggaagaatggagatagctcttgcgcgtgtcaaggtgaagcgg	480
Db	421	AATCAACGAATGAAGAAGCGGTGATTGGAAAGGGCTGCTGAGGGCAATCAAAATCAAGCCA	480
QY	481	ggggatttttctatggtccaaagcggtgactctgtcatcgatttgaaaaagaattcttgc	540
Db	481	GGAATTTCTTTATGTATGCCAAGCGGTACACTCCATGCTTTATATATGAAGGAACCTGTGC	540
QY	541	tgggagacgcagcagaactcagacaacaactacagattatagattatgacgcgaaagaat	600
Db	541	CTTGAATCATCACAAACACTCTGATACCAACATATGCGGTATACGATTTATGACCCCTTATAT	600
QY	601	gcagaagacaaagctgcgcgagcttcatctctgaaaaagaacatgaagtgaagaaggtcccg	660
Db	601	GACCAAGGCCAATAAAGAACTCTTCATATAGAAAAGCAATGGAAAGTCATATACGATACGG	660
QY	661	tcatcttcagaaacgcatagagttcacacatgaaacaaatataagatttgcctacaagaca	720

```

Db 661 CATATGATATAATGATACACCGGAAGTAAGAAAGTTGTACGCTGATCATTTGT 720
OY 721 ttagtgaatgagctactctttagtgaggaaatggaacttcaagatcagaagctta 780
Db 721 TATGTGCAATCAATATATTTCTAGTACAAATAGGAATAGGAGCCGACTCTTTT 780
OY 781 aagcagaacaaacattcctcttaccagtgtagtgaaggaggccgatagtctct 840
Db 781 CCTTCATATCAAAACCTATTTGCTGGGAGGTCTGAGCGGATCAGAGCAATCATTAAT 840
OY 841 ggtgaagatctcatctcttcaaaaaagagatcatatgtctctcgttaagcttga 900
Db 841 AATGTATTACATATGATATGATGCAATGCAAGCTCACACTTATTTCTGCTCCGATTTGA 900
OY 901 gaatttaactcgaagatgatacgaatgatactcgtctccatc 943
Db 901 GAATTTACATAGAAAGAACATGTGAATTCATGATATCTCATC 943

RESULT 6
AX006203 948 bp DNA PAT 24-AUG-2000
LOCUS Sequence 4 from Patent WO9904016.
ACCESSION AX006203
VERSION AX006203.1 GI:9929072
KEYWORDS
ORGANISM Bacillus subtilis.
SOURCE Bacillus subtilis.
REFERENCE 1. (bases 1 to 948)
AUTHORS Bacteria; Firmicutes; Bacillus/Clostridium group;
TITLE Bacillus/Staphylococcus group; Bacillus.
JOURNLS Estell, D.A.
GENEMOR Proteases from gram-positive organisms
PATENT Patent: WO 9904016-A 4 28-JUN-1999: ESTELL DAVID A (US)
FEATURES
SOURCE Location/Qualifiers
1. 948
/db_xref="taxon:1423"
BASE COUNT 286 a 192 c 244 g 226 t
ORIGIN
Query Match 33.0%; Score 311.8; DB 9; Length 948;
Best Local Similarity 58.8%; Pred. No. 1.7e-75;
Matches 557; Conservative 0; Mismatches 387; Indels 3; Gaps 1;

```

```

Db 361 ATGACTGTATAGGAAGAACGAGAAATCATTTACGGGCAATACGGCCCGCTCAAAAACGGA 420
OY 418 ctactaccatgataagcgttgagaatgagatgagctcttgcgcggtgaagtaag 477
Db 421 CTGTGTCAATGATCAACACGGGTGACTGAGAGGCTCTGCTCGAAGAAATCAAAATTAA 480
OY 478 ccggggatcttctctatgttgcgaagcggtaactgttcatbgtcgtatgaaagaaattct 537
Db 481 CCGGTGATTTCTATTATATGTCGCGAGCGAGACGTCGACGCTTGTGCAAGGGGCCCTT 540
OY 538 gctttgagacgcgcagagaaatcagacacacactacagatatatgatattagaccga 597
Db 541 GTTTTAGAGCTCAGCAAAATTCAGATGCCACATACCGGGTGTACGATTATAGCGCTT 600
OY 598 gatgcagaagcgaagcttgcgagcttcatctgaaaaagagatggaatgataagctc 657
Db 601 GATAGCAACGGAAGTCCGAGAGACTTCAATTTTGGCCAAAGCCGTCATATGCGCCAGGTT 660
OY 658 ccgtctattcagaagcagatcagttcacatgacaaatlgagattgttacaag 717
Db 661 CCCCATGTGACGCGGTATATAGATGAATGACGACGAATCAAGAAAGAAATACCATTA 720
OY 718 acattgattgaatgccttacttcttggtgggaaatlgaaacttcaagatcagaagc 777
Db 721 ACATTTGTCCAGGGAATATTTTTCGTTTATTAATGGAATGCAATGCGCAAGCTGA 780
OY 778 ttaagcagcaaaacattcctcttcatcagtgtagtgaaggaggccgtatgctc 837
Db 781 ATGGCTAGAGATGAATCCCTTTCGATTTGCAAGCTGATAGGAAGAGCGGTTGCTCAAG 840
OY 838 tctgtgagatgtctcatcttcaaaaaagagatcatatgtctgccttaccgtctc 897
Db 841 TATAGGCAAAACATGATCCGCTCAAAAAGGATGATCACTTATTTTGGCGCTCAAAATG 900
OY 898 ggaagaatttaactcgaagatatgcaaatgatactgacttccactc 944
Db 901 CCGGATTTACATTAAGAAAGAACTTGTACCTTATGCTGTCTCATAT 947

RESULT 7
BSU02562 3962 bp DNA BCT 03-JUN-1995
LOCUS Bacillus subtilis N-acetylglucosaminidase (1ytd) gene, complete
DEFINITION cds.
ACCESSION 002562
VERSION 002562.1 GI:476091
KEYWORDS
ORGANISM Bacillus subtilis.
SOURCE Bacillus subtilis.
REFERENCE 1. (bases 1 to 3962)
AUTHORS Bacteria; Firmicutes; Bacillus/Clostridium group;
TITLE Bacillus/Staphylococcus group; Bacillus.
JOURNLS Margot, P., Maue, C., and Karamata, D.
GENEMOR The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168
PATENT Mol. Microbiol. 12 (4), 535-545 (1994)
FEATURES
SOURCE Location/Qualifiers
1. 3962
/db_xref="taxon:1423"
/db_xref="taxon:1423"
/map="310 degrees"
3. 8
/feature="sigmaa"
30. 35
/feature="sigmaa"

```


Db 89617 CCCGATTTTACAGATAAAGCACTGTACCTTATCGTGTCTCATAT 89571

RESULT 10

AB011838 15012 bp DNA BCT 24-FEB-1999
LOCUS Bacillus halodurans C-125 genomic DNA, 6A fragment, clone ALBAC004.
DEFINITION
AB011838
ACCESSION
AB011838.1 GI:4512380
VERSION
KEYWORDS
SOURCE

ORGANISM

Bacillus halodurans (strain: C-125) DNA, clone_11b: lambda no. 4
clone: ALBAC004.
Bacillus halodurans
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.

REFERENCE

AUTHORS

1 (sites)

TITLE

Takami, H., Nakasone, K., Ogasawara, N., Hirama, C., Nakamura, Y.,
Masui, N., Fujii, E., Takaki, Y., Inoue, A. and Horikoshi, K.
Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125
Extremophiles 3 (1), 29-34 (1999)

JOURNAL

2 (bases 1 to 15012)

MEDLINE

Takami, H. and Inoue, A.
Direct Submission

AUTHORS

Submitted (04-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideto
Takami, Japan Marine Science and Technology Center, Deep-sea
Microorganisms Research Group; 2-15 Matsushima, Yokosuka, Kanagawa
237, Japan (E-mail: takamine@jamstec.go.jp, Tel. 81-468-67-3895,
Fax: 81-468-66-6364)

COMMENT

Sequence updated (28-Apr-1998).

FEATURES

Location/Qualifiers

1..15012

/organism="Bacillus halodurans"

/strain="C-125"

/db_xref="taxon:86655"

/clone="ALBAC004"

/clone_11b="lambda no. 4"

/map="6A"

/note="alkaliphilic Bacillus"

4..9

/note="SD sequence"

21..1421

/gene="mmgE"

21..1421

/note="similar to B. subtilis mmgE gene(74%-identity)"

/codon_start=1

/transl_table=1

/protein_id="BAA75344.1"

/db_xref="GI:4512381"

/translation="MLEQIADYVLHHEITSEAFSTAYVLLIDITGCGIILALNYPCT
KHLGPIVGTIVPRGTVPSTYLDVVGQAFNIGTAVRWLDYNDTWLAEMQSPSN
LGGIADVADFTISQENLSKGRPLKVKDYLEMMIKAHBIOGLALENSLAVGDLHYL
VKIATVATVTKLGDROETINLSNMIIDNSIRTYROAPNTGSRKSKSAAGATSA
VRLAMMAIKGEMGYOTALSAPGQFOVLFPKOKLIVNOPLDYVMENYLFKVSFPE
FHAQTAALAEIEHBEVKGRADDEITITSHSAIRIIDKRGPLYAPARDICLOYI
TAVGILGATVAEYDEDAADRIDIDLRKMVTEKKDYTKDYIDPKRSINAYVD
HKQGTSTANTEROYPLGHRREARPKIVYEKAANLGHYAKKOLEAVSYDID
RLAAMNVETELFEVK"

1425..1430

/note="SD sequence"

1439..2341

/gene="yq10"

1439..2341

/note="similar to B. subtilis yq10 gene(68-77% identity)"

/codon_start=1

/transl_table=1

/protein_id="BAA75345.1"

/db_xref="GI:4512382"

/translation="MAWIVEQPKSQKELADRFROLMKEEAIILOIPGAHDAALVAK
AGFSALYISGDAYTASRGLPDGLGIVTSTEVADRAKDLVRATNLPVLVDITGGVAIN

RBS

1856..1863
/gene="yq10"

gene

2939..3196

/note="SD sequence"

2939..3196

/gene="ydhN"

2939..3196

/note="similar to B. subtilis ydhN gene(59%-identity)"

/codon_start=1

/transl_table=1

/protein_id="BAA75346.1"

/db_xref="GI:4512383"

/translation="MKTEMAPADVPSIDKLVIHISNSNSYSHDVNRQKIOTSLSKSI
DILKIDVYRRVKKALYHKKAGDCNERKDFIGLTVRECOQV"

3446..3451

/note="SD sequence"

3458..5105

/gene="ydhO"

3458..3787

/note="similar to B. subtilis ydhO gene(69%-identity)"

/codon_start=1

/transl_table=1

/protein_id="BAA75347.1"

/db_xref="GI:4512384"

/translation="MEADFDOLSSDQIPFMIILHSGRSKILQALREYREGKEA
NDLVAERKDLNVAHEHFKLYQCATGTEKTELILLMHARDHLSLTITMKDLYKELL
ELFKMKK"

3799..3804

/gene="ydhO"

4222..4226

/note="SD sequence"

4239..5105

/note="SD sequence"

4239..5105

/note="similar to B. subtilis ydhO gene(81%-identity)"

/codon_start=1

/transl_table=1

/protein_id="BAA75348.1"

/db_xref="GI:4512385"

/translation="MNTNPLSAEYRRTVQNNITKMPQVPPVAAKSPALPAVLT
LTVFLANVFSQLEHTNMHDYIQAVQAPLVGSSGIFPLLIKIFLIQIMFGLBG
OYIINSVMDPIWNTLMIENLEAYTNGEVPNIISKPFLEITVAGSGTMTLAVFAI
LLEPKSQDKQVKGIGIGPGFVNEPIIFELPTVNMPLIIPWILSPMTVTLVTFIS
NASGLVPPPTGVAVPMTVPFPIINGIMATNSIAGGLQVLNLMIVEMTPELKSMDR
NLKKEQERSKA"

5104..5111

/note="SD sequence"

5113..6474

/gene="ydhP"

5113..6474

/note="similar to B. subtilis ydhP gene(80%-identity)"

/codon_start=1

/transl_table=1

/protein_id="BAA75349.1"

/db_xref="GI:4512386"

/translation="MPCMTKDTVTYRPEPGFMWSATSAQTGATRRDGRKNIMDH
WTEVPSRPFQGVGPODTSQFYERYKEDIQMKKEVGNHSFSSISMSRLIPDGEVN
PKGEVFNVNVIDELGOGIEPEVNLPHEDMDEMLOKIGMANREVDVAYEAYATPE
RFGDRVKKMFTHNEPIVEEGYLYDFHPNEVDPEKAIQVGYHTVLSAKAIRAYG
LNOGKGIGIILNIPSPYRSNHPADDAKAKIADAFFRSRSLVGVKSPPEELVQILK
EEGEMPIIEEDDLEIIRHNVHDIQYNYQRPVAKETLPPNPAPRPPRYEDHYAM
BGRKNVVRCHMEIYKGIYDILINUKETGIDPECTISDNKGVGGEKPFDENGVIND
DRIEFIQDHLKWHRAIQEGSNVKGTHLMTEDNMSMANVAKNKLWVGVCQFROENG
IDN"

6505..6509

/note="SD sequence"

MEDLINE REFERENCE AUTHORS	99184645 2 (sites) Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y., Masui,N., Fujii,F., Takaki,Y., Inoue,A. and Horikoshi,K. Sequencing of three lambda clones from the genome of alkaliphilic Bacillus sp. strain C-125	gene	/note="alkaliphile"
JOURNAL MEDLINE REFERENCE AUTHORS	99184646 3 (sites) Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and Horikoshi,K. Sequence analysis of a 32-kb region including the major ribosomal protein gene clusters from alkaliphilic Bacillus sp. strain C-125	CDS	/gene="BH3777" 101..652 /gene="BH3777" /codon_start=1 /transl_table=11 /product="pyrazinamidase/nicotinamidase"
JOURNAL MEDLINE REFERENCE AUTHORS	99209008 4 (sites) Takami,H., Masui,N., Nakasone,K. and Horikoshi,K. Replication origin region of the chromosome of alkaliphilic Bacillus halodurans C-125	gene	/protein_id="BAB07496.1" /db_xref="GI:10176402" /translation="MSKALIVIDYTNDFIADGALTCGCEMERPOAQGLNERPAQVYTGILGP OHVYVATDIIHADVDVHPKMLPPNINLESGSRHLVLSLOEPFESILPPVMTD KTRYSAPAGTDIHLDRHRIQLERHILHVGCTDICIHLTAVDANKGFSLVIERGVAS FNEVGHKWLGHFKSSIGADVRS" complement(846..1301) /gene="BH3778" complement(846..1301) /gene="BH3778" /note="BH3778 unknown"
JOURNAL MEDLINE REFERENCE AUTHORS	99356711 5 (sites) Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G., Sasaki,R., Hirama,C., Fujii,F. and Masui,N. Genetic analysis of the chromosome of alkaliphilic Bacillus halodurans C-125	CDS	/codon_start=1 /transl_table=11 /protein_id="BAB07497.1" /db_xref="GI:10176403" /translation="MRGWIAICVVALTCLVGCHEMERPOAQGLNERPAQVYTGILGP GPANYDNVYKRSIAPDAREONSGSFPTATETPTIATIDETIRTSIQDEHLRSMV MTASSHAWVNVPEPRNMKKEQRNKIMNDLREDMQROIPIROYVHLSQKG" complement(1382..1996) /gene="tdk" complement(1382..1996) /gene="tdk" /EC_number="2.7.1.21" /note="BH3779" /codon_start=1 /transl_table=11 /product="thymidine kinase" /protein_id="BAB07498.1" /db_xref="GI:10176404" /translation="MQMIRKEGWIEVYICSGMSFKSEELIRRYRATYRLKVOYFKP AIDNRYSEQVNSHNGKVCALPVERKMSITLDEGSDTDIVLDEQFDEVTGIVD VLADGVAVICAGIDQDFRGEPEPTPTLMAKVEYTKQALICPVCGSPASRTRLID QKPSYHDPITLIVGAALAEYPRCHCHQVYNNKSLPEASGE" complement(2159..2404) /gene="rpmB" complement(2159..2404) /gene="rpmB" /note="BH3780" /codon_start=1 /transl_table=11 /product="50S ribosomal protein L31" /protein_id="BAB07499.1" /db_xref="GI:10176405" /translation="MKRGCIHPDYKVKVYFMDSTGFEKLSGSTKTSNETIEMDGTNP LIVKISSDSHPFTGKQKLADAGKRVKRTNKL" complement(2516..3787) /gene="rhc" complement(2516..3787) /gene="rhc" complement(2516..3787) /note="BH3781" /codon_start=1 /transl_table=11 /product="transcriptional terminator" /protein_id="BAB07500.1" /db_xref="GI:10176406" /translation="MGVHSELNMMKIRLEYELARKYKTSYSLKRLKELPAIILKGO AEEGLAFMEGVLEIISGSGFGLRPIINWPSSEDIYISASQIRREDLFGNCGVSGKY RPPENRERHGLHVEVKNEDDPTSKERHFPALPLPERKUNLETTPGHISRSII DMISPVGGGGLVAPRKAGKTSLSKELNSIYNHHPVDGLILDLRPEEVTDIE RSVADAVYSGTDEVPENHLKVALERAMRIVERKQVYIILMDSITRLAVYNDVI PESGRTLSGGIDPAARPRKPRFGAANNIEGSLITLIALVEGSRMDADVITYEFK GTGMWELHLDRKLSERRIPALIDIRRSSTKRELLMKSLKLELMAVARTNMDSEFV DHFRVRYETKTNDPEEAMKDMTNGTRKK" 3913..4257
JOURNAL MEDLINE REFERENCE AUTHORS	20263314 10 (sites) Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T., Hirama,C., Fujii,F. and Takami,H. Characterization and comparative study of the rrm operons of alkaliphilic Bacillus halodurans C-125	gene	/note="alkaliphile"
JOURNAL MEDLINE REFERENCE AUTHORS	20426005 11 (bases 1 to 305153) Takami,H. and Takaki,Y. Direct Submission Submitted (22-MAR-2000) Hideyo Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group, 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan (E-mail: takami@jamstec.go.jp) URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSAR/Research.html, Tel.81-468-67-3853, Fax:81-468-66-6364)	CDS	/location="Qualifiers 1..305153 /db_xref="taxon:86665"
FEATURES Source	1..305153 Location="Qualifiers 1..305153 /db_xref="taxon:86665"	gene	

gene
/codon_start-11
/transl_table-1
/product-"conserved hypothetical protein"
/protein_id="AAK04859.1"
/db_xref="GI:12723679"
/translacion="NNITTPDLFGFVTLVFMIGVRIINERPKMTNPILFLFAVTV
ILILSLALPKYKNYKSGISLINDLIGPSTVALGIPKIFELMKHARSIVSIALTA
LVNCITFLACAFKPSYKSLASISIPKSVTTAAAGMISOKMGVSDIYVVVYVNGTIL
TVYLGGLPLKLRITRDPAQAIGALGCGHVGSTALELKGQGMALISGYGIMK
VIRAPILAKITLIG"
complement(785..1189)
/gene="yhnd"
/note="1172519"
complement(785..1189)
/gene="yhnd"
/note="EVIDENCE BY HOMOLOGOY BIO15.02 HYPOTHEICAL.
Conserved. 50% identical to PH1801 of *Pyrococcus
hottkoshii*."
/codon_start-11
/transl_table-11
/product-"conserved hypothetical protein"
/protein_id="AAK04860.1"
/db_xref="GI:12723680"
/translacion="MKIYLQILFIIGFSEIGNVSNFRLPVGSLIGLILFLALOF
KLEIFHNVHVASGFLINMTTILFLPAGGIMAKMNILSHWAQILIVVALLINMIL
LGLKEMIVKYEKGVYALKEPTVYSTEDQ"
1467..1935
/gene="yhne"
/note="1173151"
1467..1935
/gene="yhne"
/note="EVIDENCE BY HOMOLOGOY BIO15.02 HYPOTHEICAL.
Conserved. 36% identical to Y1XS of *B. subtilis*."
/codon_start-11
/transl_table-11
/product-"conserved hypothetical protein"
/protein_id="AAK04861.1"
/db_xref="GI:12723681"
/translacion="MNDVVKIYEDPFIPIPHIOPPELPAVMEKMGQDMKLSTVYDKES
GIEIDFASLISLIDLTISPPPEKGLLEVASRPAEPLIKRASFAGANGPYTIL
PVTLKXINKKEITDGLVSFDEGLIVYDLTRIKNIEIPLSAVAKQIKVNF"
2100..3248
/gene="nusa"
/note="1173813"
2100..3248
/gene="nusa"
/note="EVIDENCE BY HOMOLOGOY BIO11.02 RNA synthesis,
modification, and DNA transcription. COULD PARTICIPATES IN
BOTH THE TERMINATION AND ANTI-TERMINATION OF TRANSCRIPTION"
/codon_start-11
/transl_table-11
/product-"transcription termination protein NUSA"
/protein_id="AAK04862.1"
/db_xref="GI:12723682"
/translacion="MSKEMLAFAMLEROGIKSEIVTEAIEQATIAAYKROYGOAGN
VKIYEPBKGNFVYSTREVVDEVDSELESDAHELPHYEIGDKIMPEEKPKDE
ARRAAGAKOVINOKKREBEERTIYNEYSRKDEIIQGVYEKADAVYVNLGRVDA
LTKDQPONFHVGGRVAKVYTVYDVLTPKGTREYISRTAPDKMLKFEKREIPEVFD
TEYKSTIANRAGBRAKVIYVOSHEDVADIGVGTGSGRSOGIIRLAGEKGLIEMW
EDKATVIAARAGRAKVIYVOSHEDVADIGVGTGSGRSOGIIRLAGEKGLIEMW
DIRAAEPNLDIPEYVGDGESEIPTTENELVVEEVEPEPSEEE"
3414..3740
/gene="ybhg"
/note="1175136"
3414..3740
/gene="ybhg"
/note="EVIDENCE BY HOMOLOGOY BIO15.01 HYPOTHEICAL. General"
/codon_start-11
/transl_table-11
/product-"HYPOTHEICAL PROTEIN"
/protein_id="AAK04863.1"
/db_xref="GI:12723683"
/translacion="KKQKTEPKRSKSLVSNDFPKDULRIAYNKGEISIPSGKAGHG

[illegible]

	gene	gdnkxenoefkerfpaadakitllvksnfpdvymkojhghepfkrvadtylvsvlldsqoc ltidknepxpevkgaahfilpslniekwnlsqleliaasnpa"
	CDS	/gene=".9517 /gene="rimg" /note="L180805" 5077..9517 /gene="rmac" /note="EVIDENCE BY HOMOLGY B1009.08 Marr-family regulators. belongs to the Marr family" /transl_table=1 /codon_start=1 /product="transcriptional regulator" /protein_id="AAK04868.1" /db_xref="GI:1273688" /translation="MKTDPDKVNMWLIQLFHEPMILEEQFLKDSAVSDVTKLEQILITL LVHFGTCRATDIAKNOKLALSTTTTLNKLDEKGYIEKRSTADRRAYTHLLSKSGIK ELCATHRREFPHRIIDNFLDSVYGTEKKLAGELATLHDLSLENMK" 517..10494 /gene="fabh"
	gene	
	Query Match	29.3%; Score 277.2; DB: 1; Length 10607;
	Best Local Similarity	57.3%; Pred. No. 7.9e-66;
	Matches 342; Conservative	0; Mismatches 398; Indels 6; Gaps 2.
Oy	I atgcagactgaacggttatttttccaagccgttggtttaagaagaatltggcgcggaac	60
Db	7708 ATTATGAGAACCAACATTGTCTTGAACATCAGATTTGGCAAAGAAAAAATCTGGCGGCAGAC	7767
Oy	61 gcttcctgatcttttggcatalacccttcgcgcacaacgaaggcgagcttgcgctt	120
Db	7768 CATTTGAAAGAGTTTGCGTATGATTGGCCATCAGACAAGTAGTGTAATTTGGCGTAT	7827
Oy	121 gcgcgcatacaaatgtgtcaaagcgttgttcaaaacggaaatgtaaaagggttcacagctc	180
Db	7828 TCGGTCATGCCATGTGTGTGCACAACTGGTATGSCGAATTTAAGTCAAAAACCT	7887
Oy	181 agcgaaattatgggaacatcacagaacattatctggacagcgttgaaggagccggttcct	240
Db	7888 GACCAAATTATGCCAAGTCACCGCAAAATGGTGGTGTAATAAAGAAGATTTTTCCC	7947
Oy	241 ctgcttcaaaaatatagatgctgcaccaaggaacttatctgtcagtgcacccgaatgat	300
Db	7948 TTACTAACTTMAATTTTAAATGATCCATATGACGCGCTTCTGTGCCAAGTTCATCCAGATGAT	8007
Oy	301 gaatattgccacatacatgaaaacgctgagccttggaaaaaacgaatbcctgtccatatt	360
Db	8008 GAATATGCAAAAAAATCATGAA--GATGAAGTTGGAAAACTGAATGTTGTCATATT	8064
Oy	361 gatgcgaaaagaatgcgagagatttatatggcccacaatgacaacagaagaagaacta	420
Db	8065 TCACGTGAAACACGCGCTGAGATTATTCATATGACATATGCTAAATCACGTGAAGAATTA	8124
Oy	421 actaacctatgtagaagcttgagaatggagatgagccttgcgcgctgtcaaaugtcaaacg	480
Db	8125 GCACAAATGATTTAAATCTGTGATGTGGATCTTTGTTCATGAAGTAAAGTGAANAACA	8184
Oy	481 gggaatcttttctaigtgcgaagcggtaactgttcaatlgaaatlggaaaagaatcttgt	540
Db	8185 GGGGATTTCTTCATATTCGTCAGAGAAATGACACGCAATCGGTGCGAATTTGTTATT	8244
Oy	541 ttgagaagcgaagaacctcagaacaaoactcagaatatatgttlaagccgaaaagat	600
Db	8245 CTTTGAACACAAACAACTTCTGATCAACTTACCGTGTGTTATGATTTGCACCGTAATAAT	8304
Oy	601 gcaagaagcgaagcttgcgcagcctcctcgaaaagagcactgaagtgatagagatgcc	660
Db	8305 GACCAAGGAATCTACGTGAATTACATATTCACAACATCAATTGATGATTAATATTCG	8364
Oy	661 tctatccagaac---gcatcaogtftaacccatgaacaatltgaggttgccttaaacag	717
Db	8365 GGCGACAAATCCCTGAAATTCATTTAAACTGAAAAATTTGCTGATGCAAAAATTTACA	8424
Oy	718 acatitgatgatgcctacttcttcggtggggaacggaactatccaagtcagaacg	777

DB	Accession	Source	Reference	Journal	Medline	Remark
DB	8425	ACNCCTGTGTAACAAATTTCTTGATGTATATTAATGCAAAATTCATGATGCATGAA	8484			
OY	778	tttaagcagcaaaaacccctctctcttcttaacgttgattgaagggagggccgtatgac	837			
DB	8485	TTTACCAAGATTCGTCGATACACTTAATTTCTGTCTACTGATGTCAGGAAATTAACA	8544			
OY	838	tctgttgatgatgctatcccttccaaaaagagatcatalatgtytgccttaagcttc	897			
DB	8545	ATTGATGGAATTAATATTCAGATTTGAAAGAGACCTCATTTTCATCTTACCAAGCAACAT	8604			
OY	898	ggagaaatttaacatcgaagatcagatcagatgtaatcgtctccacc	943			
DB	8605	GAAAAATGGAATTTGTCTGTGTCATTAAGAAATTAATGTCAGCAATC	8650			
RESULT 13	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
STREP1	1379 bp	DNA	BCT	04-FEB-1999		
STREP1		S. mutans pml gene for mannosephosphate isomerase (complete cds)				
SCG		scg gene for fructokinase (partial cds).				
D16594.1	GI:451214					
fructokinase		mannose-6-phosphate isomerase; mannosephosphate isomerase.				
Streptococcus mutans (strain GS-5)		DNA.				
Streptococcus mutans		Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.				
1 (bases 1 to 1379)						
Sato, Y.		Direct Submission				
Submitted (05-JUL-1993)		to the DDBJ/EMBL/Genbank databases.				
Sato, Y., Yamamoto, Y., Kizaki, H. and Kuramitsu, H. K.		Isolation and sequence analysis of the pml gene encoding phosphomannose isomerase of Streptococcus mutans				
FEMS Microbiol. Lett. 114 (1), 61-66 (1993)						
94123951		Erratum: [[published erratum appears in FEMS Microbiol Lett 1994 Aug 1;112(1):130]]				
3 (sites)						
Sato, Y., Yamamoto, Y., Kizaki, H. and Kuramitsu, H. K.		Isolation, characterization and sequence analysis of the scrk gene encoding fructokinase of Streptococcus mutans				
J. Gen. Microbiol. 139 (Pt 5), 921-927 (1993)						
93329360		Submitted (05-JUL-1993) to DDBJ by:				
Yutaka Sato		Department of Biochemistry				
Tokyo Dental College		1-2-2 Masago, Mihama-ku				
Chiba, Chiba 261						
Japan		Phone: 043-279-2222 x2703				
Fax: 043-279-2052.						
Location/Qualifiers						
1. .1379						
/organism="Streptococcus mutans"						
/strain="GS-5"						
/db_xref="taxon:1309"						
1. .72						
/gene="scrk"						
<1. .72						
/gene="scrk"						
/RC.number="2.7.1.4"						
/note="3, end of scrk"						
/codon_start=1						
/transl_table=11						
/product="Fructokinase"						

```

/protein_id="BA04020.1"
/db_xref="GI:45125"
/translation="PAAVENSATLGLAKKIAAR"
RBS
230..234
/note="ribosome binding site"
gene
241..1191
/feature="pm1"
241..1191
/feature="pm1"
CDS
/EC_number="5.3.1.8"
/codon_start=1
/transl_table=1
/product="Mannosephosphate isomerase"
/protein_id="BA04021.1"
/db_xref="GI:45126"
/translation="MEGPIPLIOSQNHKKIMGNLRKEFGYDIPEFTGEVNAISAP
NGSVYKGVYKGVPLDELYAHEHLEFGNSKSSVEPLITLIDANMISVQAPDNAY
ALEHGEIGKECEKYVISADEGAETIYGHAKSKKEELROMLADMDHILITKPYAG
DEPYRSCYTHMISRGIMILETQOSSEPTVRYVDPKRDQCKRRLHIEQSIDVYLI
CKRANATPAMISLDGLETVLVSPFTVYKMOISGVKRAQTAFTLVSVLAGQERI
TGEQYALRKGDHLIPNTIKSMQFGDLEITASHNEC"
stem_loop
1248..1289
/function="putative transcription terminator"
-35_signal
1309..1314
-10_signal
1332..1337
BASE COUNT 438 a 215 c 300 g 426 t
ORIGIN
Query Match 28.1%; Score 265.8; DB 3; Length 1379;
Best Local Similarity 57.4%; Pred. No. 9,2e-63;
Matches 539; Conservative 0; Mismatches 392; Indels 8; Gaps 3;
6 gactgaccgctatttttcaagccgctttcaaaagaattggggggaccgctt 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 GAAGGACCTTTGTTTACAAATGCAATATAAAATCTGGGGGCAATCGGCTC 303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66 ag-ctgatttggcattccattccgcaacagcaggagctgctggc 123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 AGAAAGATTGGTTTATCATCTCTGCTGAGACGAGAGAAATTTGGGCAATTC 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 ggcacataatgctcaaaagcttgcataaagcgaatgataagggttcaagctcagc 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 GCCCATCTTAATGAGAGTCTGTGTTTAAATGAGTGTAAAGGTCCTTGAT 423
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 gaattatggaaacatcacacacattatcggacagcttgaaggagaccgttccctc 243
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 GAATTATGACAGACATGAGACATGTCGGCAATACCAAGATCTGTTTCCGCTT 483
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 ctacaaaataatgatcgtcagcagcactatctgttcagtgctcgcacacgaatgata 303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 TTAACTAAATCTTGAATGCCAATGCGTTAGTTCATCAAGTTCAATCCGATATGCT 543
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 tatggcaacatcagaaagcgttgagtgaaagaaagcgtgtacattatgat 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 TAGGCTTTGGAACATGAA--GATGAGTAGAAGACAGAAATGATGATGATTCATCA 600
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 tgcacaaagatgcagagatttattatggccaacatgcaacacaaagaaactact 423
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
601 GCTGATGAAGAGGCTGAAATTAATCTATGACACGAAAGTAAATGAATTCGCT 660
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 accatataagacgttggagaaatgagctcttgcgcctgtgaaaggtaaagccgg 483
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 CAATGATTTGCGCGGGGATGGGATCACCTTTTGACAAAGATTCCTGTTAAAGCTGG 720
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 gatttttctatgtcgaagcgttactgttcagtgatgtaaaagaaatcttgcctt 543
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 GATTCTTCTATGTTCTCAGTGGACATGATGATGATGATGATGATGATGATGAT 780
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 gagaagcagcgaactcagacacacacacatataatgataagcgaagaaatgca 603
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 GAGACACACGACGATCAAGTATGACGACCTACCGCGTTATGATTTGATCGTAAAGATGAT 840
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

604 gaaggcaagcgtgcgagcttcacatcgaaaaagacatcgaatgataagagctccgct 663
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 CAAGACCAAAACGAGCTTCACATGACATGCAATGATGATGATGATGATGATGAT 897
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 attcagaagcgcatacagcttcaccatgaacaaatgagagattgttcaacagacatg 723
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
898 AAGCCGCAATATGCGACCGCAGCTTGGTTATCTTACAAAGCGCTTGAAACAGGCTTG 957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 attgaatgcctacttccttgcgtgggaaatgagactatcagatcagaacagcttaa 783
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
958 GTTCCAGCCCTTCTTACTGCTATTAATATGCGAGATGATGATGATGATGATGATGAT 1017
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 cagcaaaaacatcctcttactatcagtgatgaaagggagggccgtatgactctgt 843
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1018 CAGAACGCTCTTACTTCTTGGTGAATGCTCTTCTGACAGAGTGTATACGCTTGGT 1077
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
844 gactgtctatccttcttcaaaaagagatcatatgttgcctcctcaggtcttgagaa 903
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1078 TTAGAACATATGCTTAAAGAAAGAGATCATCTTATCTCTTAACTATTAATCAT 1137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904 ttaaactgaaagatcagaaatgatacgtctccat 942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1138 TGGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 14
AC027136 49897 bp DNA HTG 30-JAN-2001
LOCUS staphylococcus aureus clone sabac-106, WORKING DRAFT SEQUENCE, 1
DEFINITION
ordered pieces.
AC027136
AC027136.5 GI:8567822
VERSION HTG: HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS staphylococcus aureus.
SOURCE staphylococcus aureus.
ORGANISM Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
1 (bases 1 to 49897)
Loh, P., Qi, S., Ray, L., Ford, B., Worrell, V., Iandolo, J. and Roe, B.A.
Staphylococcus aureus BAC Clone sabac-106
Unpublished
2 (bases 1 to 49897)
Loh, P., Qi, S., Ray, L., Ford, B., Iandolo, J. and Roe, B.A.
Direct Submission
Submitted (28-MAR-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of 'N'. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 49897: contig of 49897 bp in length.
FEATURES
location/Qualifiers
source
1..49897
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
/clone="sabac-106"
/clone_lib="OU staph library"
BASE COUNT 15753 a 9237 c 7267 g 17640 t
ORIGIN
Query Match 26.7%; Score 252.2; DB 71; Length 49897;
Best Local Similarity 55.1%; Pred. No. 7.6e-59;
Matches 516; Conservative 0; Mismatches 418; Indels 3; Gaps 1;
7 actgacacgtatttttcaagcgttcttcaagaagaattggggaccgcttca 66

```



```

Db 35576 ACTATGCGATTATTTTACACCCGTTTAAAGACAGATTGGGGCGCATCATTA 35535
QY 67 gctgatttggctataccattccgtacacaagaagggagtgctggcgttttggcgcg 126
Db 35636 AAGGCAATTAACATATGATATGCTATATGAAACAACTGGGAATGCTGGCAATTTCTCA 35695
QY 127 catcaaatatgtaaaagcgttgcataaaggaatgtaaaaggggttcacgctcagcgaa 186
Db 35696 CATCCAAATGCTCTATATACATATGATATGCTCATACAGATATGACCTTGAATCA 35755
QY 187 ttatggacaatcacagaacattatctggacagcgttgaagggacgcttccctcctt 246
Db 35756 CTTTGGTCAACAAATCGTGAATTTATTCGATTAATGATTCACGAGATATGTTCCCTCGCTA 35815
QY 247 acaaaatataatgatactgacagagactatctgtcaggtgacgacgacgaatgaat 306
Db 35816 ACTAAATATTAATGATCCCAATGATTAATATCTGTCAAGTTCACCAAGATGATGATAC 35875
QY 307 gccaaatatacaaaacggtgagcgttgaagaaacagaaatgctgtatcatattgattgc 366
Db 35876 GCTTTAAACACGAA---GGCGAAGTACAGTAAACAGAAATGTTGATATTTTATGATCA 35932
QY 367 caaaagatcgcgaattattatgagcacaatgcacaacaagaaggaactaactacc 426
Db 35932 GCGCCAGGCTGTAATTAATTAATGCGTACATGCAATATTAACAGACGTTAATTAAC 35992
QY 427 atgataagcgtggaagatggaatgagcctctgcgcgtgtaaaaggaagcggggagat 486
Db 35992 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 36052
QY 487 ttctcattgctgcaagcgtgactgctacatgcatgcatggaaggaactcctgcttggag 546
Db 36052 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 36112
QY 547 acgtagagagaactcagaacaactacagatatatgattatgacccgaagaatgagcaa 606
Db 36112 ACACAAATATCTCTCGATGAGATACCGTATTTATGACTATGATAGACAGATTAATTAAT 36172
QY 607 ggcagctgcgcgaagcttcaatcgaagaaaggaatggaatgagagagccgctcatt 666
Db 36172 GGTCAATTTGCGTATTTGCTTACAGTAAAGTAAAGATGATTAATTAATTAATTAATTAAT 36232
QY 667 ccagaaacgcatcaatcctcacaatgcaaatggaatggaatggaatggaatggaat 726
Db 36232 GAACCAATATCTCTCTCTATTAACACTCAATGACACACACACACACACACACACACACAC 36292
QY 727 gaatgcgtactcttgcgtggaaggaatggaatggaatggaatggaatggaatggaat 786
Db 36292 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 36352
QY 787 caaaacacatcctccttcaatcagtgatggaaggaagggcgatagatcctgagag 846
Db 36352 CCACATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 36412
QY 847 tatctcattcctcacaagaaggaatgcatatgctgctcctcctcctcctcctcctcct 906
Db 36412 AAGCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 36472
QY 907 aaactcgaaggaatggaatggaatggaatggaatggaatggaatggaatggaatggaat 943
Db 36472 CACTTTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 36509

```

```

ORGANISM      Staphylococcus epidermidis
REFERENCE      Bacillus/Staphylococcus group: Staphylococcus.
AUTHORS        Kimmerly, W.J., Taylor, J. David, Nelsen, A.J., Godlevski, M.M.,
                Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
                Lisenbee, S., Ashanti, C., Altschuler, G., Mam, L., Shepherd, N.S.,
                Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
                Furdon, P.J.
TITLE          Transposon-mediated sequencing of the Staphylococcus epidermidis
                genome
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 2815)
AUTHORS        Taylor, J. David, Kimmerly, W.J., Nelsen, A.J., Godlevski, M.M.,
                Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
                Lisenbee, S., Ashanti, C., Altschuler, G., Mam, L., Shepherd, N.S.,
                Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
                Furdon, P.J.
TITLE          Direct Submission
JOURNAL        Submitted (22-MAY-2000) Departments of Genomic Sciences and
                Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
                Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES       location/Qualifiers
                1..2815
                /organism="Staphylococcus epidermidis"
                /strain="SRI"
                /db_xref="taxon:1282"
                /clone="step.1023d05"
                /base="518 c 315 g 1024 t"
BASE COUNT    958 a 518 c 315 g 1024 t
ORIGIN
Query Match      24.0%; Score 227; DB 2: Length 2815;
Best Local Similarity 52.9%; Pred. No. 5,3e-53;
Matches 488; Conservative 0; Mismatches 435; Indels 0; Gaps 0;
QY 13 cgttatcttccaagcctgtttccaagaagaattggggcgagaccgttaactgat 72
Db 1404 CCGTATTTTAAACCCATTTTCTGATTAAGTATGCGGCACTGATTAATTCGTCGA 1345
QY 73 ttggtatcacatcctcgtccacaagaaggaagtgctggcttggcgtcgcgcacaa 132
Db 1344 TTTGGGTATCAACTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1285
QY 133 aatggtcaagcgtgtgtccaagaaggaatgataaggggttcacgctcagcgaattatg 192
Db 1284 CACGGAATAAGTGTATGAAATGATATTTGCTGCTGCTCAACATGATGATCAAGATATG 1225
QY 193 gaacatcaagacattatctggaacgcttgaagggagcgtttccctcgtctacaataa 252
Db 1224 AACAAATCAATAGAAATTTTGGAGATTTCCAAAGTAAAGATTTCCATTAATTAATTAAT 1165
QY 253 atattagatcgtacccaggaactatcgttcaagtgatcgcgaatgataatgccaac 312
Db 1164 ATTGTATGATGCTGCTGCGCATATTTCTATTAATTAATTAATTAATTAATTAATTAATTAAT 1105
QY 313 atacatgaagaacggtgagcttggaaaacaagaatgctgtacatctattgttgccaaaaa 372
Db 1104 GAACAGAAAGAGTCAATATGAAATCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1045
QY 373 gatgcgagatattatgagcacaatgcaacaagaaggaaggaactacacatgata 432
Db 1044 GGTGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 965
QY 433 gagcgtggaagatggaatgagccttgcgcgtgtaaaaggaagccggggatcttttc 492
Db 984 GAGCAAGGTATCTTTGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 925
QY 493 tatgtccaagcgtgacatcttcattgcaatgcaatgcaatgcaatgcaatgcaatgcaat 552
Db 924 TTTATCAAGAGTGAAGATCAATTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 865
QY 553 cagaactcgaacaacacacagatataatgattatgacgaagaagtgcaagaagcgaag 612

```

```

RESULT 15
AF269779/c 2815 bp DNA BCT 01-AUG-2000
LOCUS      Staphylococcus epidermidis strain SRI clone step.1023d05 genomic
DEFINITION
sequence.
AF269779
VERSION    AF269779.1 GI:9623676
KEYWORDS   Staphylococcus epidermidis.
SOURCE

```

```

Db      864  CAAATCATCGAATATTCATATGAAATTTATGATATCATAGAAAACTGATAATAGTGAG 805
      613  CTGCGGAGCTCCTCAGCTGGAAGAGCATGAGTGAAGAGGCCGCTCCTATCCAGAA 672
      804  GAACGTGAATTAATATAGATGAAGCATTAAGTGTATTAATTAATTCAAATGAACCTA 745
      673  CGGCATACAGTTCACCATGACAAATGAGATGCTGCTACAGACATGATGAAATGC 732
      744  AATATCATCTCTCAAAATGAACTGATAGAAATCAATTTGATACATATGTATAGT 685
      733  GCTTACTTTCGCTGGGGAATGGAATCAGATCAGCAAGCTTAAGCAAGAAAA 792
      684  GATTTTCTTACTATGATGATAGGATATTTCTGCTACTCTAAATTAATGAAAGCTTGA 625
      793  CCAATCCTCTCTATCAGTGTGATGAGGAGGCGTATGATCCTCGTGTAGTATGTC 852
      624  GAATTTTGTCTGTGTTCTGTTTATGATGACAAAGTAACCTAATGTAGATGATATATA 565
      853  TATCCTTCAAAAGAGGAGATCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 912
      564  TATGATATATCTAAAGTTCAAAGTTCGATTAATCTCTGAAGATTTAGATAGTGTTC 505
      913  GAAAGGATATGCAAGATGATGAT 935
      504  GAAAGGATTTTAACTATATCAT 482

```

Search completed: May 17, 2001, 15:38:04
 Job time: 11435 sec

Fri May 18 10:38:00 2001

us-09-462-846-2.p2n.rge

Page 1

OK of: US-09-462-846-2 to: GenEmbl:* out-format : pfs

Date: May 17, 2001 4:11 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL-frim-p2n.model -DEV-xip  
-O/cgml2/usrt0.spool/us09462846/runat_17052001_115428_937/app-query.fasta.1.376  
-DB-GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000  
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462846.ecgml_1.5014 -NCPV=6 -ICPV=3 -LONGLOG -NO_XLPHY  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-462-846-2  
Query length: 315  
Database sequences: 1283235  
Database length: 1216004940  
Search time (sec): 1159.440000
```

score list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_pati:AX006200	+ 1692.00	2627.47	3.9e-138	945	! AX006200 Sequence 1 from Patent
gb_ba2:BSU0007	+ 1692.00	2574.91	216750	1	! D291110 Bacillus subtilis com
gb_ba1:AB011838	+ 988.50	1505.20	1.3e-75	15012	! AB011838 Bacillus halodurans
gb_ba2:AP001520	+ 988.50	1476.08	5.3e-74	30513	! AP001520 Bacillus subtilis N-ace
gb_pati:AX006203	+ 978.50	1516.56	3.0e-76	948	! AX006203 Sequence 4 from Patent
gb_ba2:BSU02562	+ 978.50	1502.05	1.8e-75	3962	! U02562 Bacillus subtilis gene
gb_ba2:BAC0019	+ 978.50	1502.05	2.5e-75	4162	! D45046 Bacillus subtilis N-ace
gb_ba2:BSU00019	+ 958.50	1449.95	1.6e-74	945	! AX006205 Sequence 6 from Patent
gb_pati:AX006205	+ 958.50	1449.95	1.5e-72	36308	! D88802 Bacillus subtilis DNA
gb_ba2:D88802	+ 958.50	1432.83	1.4e-71	213190	! D29107 Bacillus subtilis com
gb_ba2:BSU00004	+ 878.50	1357.01	2.3e-67	1379	! D16594 S.mutans pmi gene for m
gb_ba3:SRP01	+ 867.00	1319.37	2.8e-65	10607	! AF006310 Lactococcus lactis s
gb_ba2:SRP01	+ 867.00	1250.34	2.0e-65	147006	! AT449826 Streptococcus pneum
gb_ba2:SRP01	+ 819.50	1230.42	2.6e-60	49897	! AC027136 Staphylococcus aureu
gb_hg12:AC027136	+ 719.00	1075.04	1.3e-51	44449	! AC027136 Staphylococcus aureu
gb_hg17:AC074320	+ 719.00	1074.00	1.3e-51	44487	! AC074320 Staphylococcus aureu
gb_ba1:AC074320	+ 719.00	1072.91	1.5e-51	55389	! AC074320 Staphylococcus aureu
gb_ba1:AC078934	+ 719.00	1072.91	1.5e-51	55390	! AC078934 Staphylococcus aureu
gb_ba2:AF269779	+ 685.00	1100.17	4.6e-53	2815	! AF269779 Staphylococcus epide
gb_pati:AF64844	+ 718.00	1051.57	2.3e-50	2111	! AF64844 Sequence 1 from Patent
gb_ba2:AF269301	+ 580.50	884.02	5.0e-41	3473	! AF269301 Staphylococcus epide
gb_ba1:AB016431	+ 478.00	723.65	4.3e-32	3754	! AB016431 Staphylococcus aureu
gb_ba1:AB016431	+ 426.00	603.58	2.1e-25	213732	! AF001862 Deinosporus radiod
gb_ba2:AF270310	+ 419.00	635.97	3.3e-27	2433	! AF270310 Staphylococcus epide
gb_ba1:AB001744	+ 391.00	571.36	1.3e-23	21343	! AB001744 Thermotoga maritima
gb_ba1:AB001955	+ 213.50	305.80	8.1e-09	6951	! AB001955 Streptococcus sanguis
gb_ba1:AB001104	+ 208.50	290.79	5.0e-08	14878	! AB001104 Archaeoglobus fulgid
gb_ba2:DLA01116	+ 166.00	227.31	5.2e-05	3883	! AB001116 Lactococcus lactis 10
gb_ba1:AB001116	+ 166.00	222.37	0.0004	40744	! AB001116 Lactococcus lactis 10
gb_ba1:AB001116	+ 163.00	245.89	1.8e-05	619	! AF015751 Lactococcus lactis 10
gb_ba1:AF015751	+ 160.00	245.89	1.8e-05	619	! AF015751 Lactococcus lactis 10
gb_pati:AX047314	+ 145.50	195.08	0.0139	11661	! AX047314 Sequence 28 from Pat
gb_pati:AX047311	+ 145.50	194.03	0.0116	12766	! AX047311 Sequence 25 from Pat
gb_ba2:ECOMANA	+ 145.00	213.32	0.0012	1604	! M1580 E.coli manA gene encodi
gb_ba1:AE000257	+ 145.00	194.77	0.0124	10914	! AE000257 Escherichia coli K12
gb_ba2:D90804	+ 145.00	191.40	0.0191	15465	! D90804 E.coli genomic DNA, Kc
gb_ba2:D90805	+ 145.00	189.30	0.0250	19211	! D90805 E.coli genomic DNA, Kc
gb_ba1:AE005385	+ 139.00	185.64	0.0401	10679	! AE005385 Escherichia coli O15
gb_ba3:MY020B11	+ 133.00	164.45	0.6063	36330	! Z95121 Mycobacterium tubercu

```
gb_p13:SCYJ1036W + 116.50 163.17 0.7148 2912 1 249311 S.cerevisiae chromos  
gb_ba3:SPM1PHI + 115.00 166.33 0.4767 1650 1 X57117 S.typhimurium pmi ge  
gb_ba1:AE004115 + 114.50 146.26 6.25 12116 1 AE004115 Vibrio cholerae c  
gb_p13:SCORF512 + 112.00 156.88 1.60 2703 1 248229 S.cerevisiae tRNA-Va  
gb_pati:AX065227 + 110.00 160.81 0.9672 1305 1 AX065227 Sequence 353 from  
seq_name: gb_pati:AX006200
```

seq_documentation_block:

LOCUS	AX006200	945 bp	DNA	PAT	24-ANG-2000
DEFINITION	Sequence 1 from Patent WO9904016.				
ACCESSION	AX006200				
VERSION	AX006200.1	GI:9929071			
KEYWORDS	Bacillus subtilis.				
SOURCE	Bacillus subtilis.				
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.				

REFERENCE 1 (bases 1 to 945)
Estell,D.A.
Proteases from gram-positive organisms

JOURNAL Patient: NO 9904016-A 1 28-JAN-1999;
GENECOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)

FEATURES
Source location/Qualifiers
1..945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"

BASE COUNT 288 a 180 c 232 g 245 t
ORIGIN

alignment_scores:
Quality: 1692.00 Length: 315
Ratio: 5.371 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-846-2 x AX006200 ..

Align seg 1/1 to: AX006200 from: 1 to: 945

```
1 MetThrThrgluProleuphethelysProvalPhelygluargl1eTr 17  
1 ATGACGACGTGACCGTATTTTCAAGCCGTTTCAAGAAAGAAATTGG 50  
17 pGlygluThralaenualasphheglytyrThrileproserglnargT 34  
1 GGGGGGAGCCGCTTACCTGATTTGGCTTACCATTCGTCACACGAA 100  
34 hrGlygluGlytTPAlaAlaAlaHisGlnasnglyGlnserValVal 50  
101 CAGGCGAGTGTGGCTTTTGGCCCGCATCAAAATGATGCAAGCGTTGTT 150  
51 GlnasnglyMetTyrlsGlypTherleuSerGlnleuTrpGlnHisH 67  
151 CAAAGCGAATGATTAAGGGGCTTACCGTCAGCAATTAAGGAAATCA 200  
67 sArgHisleupheglyGlnleuglyGlyAspArgpPrepoluThrl 84  
201 CAGCATTTATTTGGACACACCTTGAAGGAGCCGTTCCCTGCTTACA 250  
84 yslleuaspaAlaaspGlnaspleuservalGlnvalHisProasnsp 100  
251 AAATTTAGATGCTGACGACGACTTATCTGTTCAAGTGATCGAATGAT 300  
101 GlutyrAlaasnIleHisGlnasnglyGluLeuglyLysThrGluGlyTr 117  
301 GAATATGCAACATACATGAAAGCGTACCTTGAAGAAAGAAATGCTG 350  
117 pTyrlleleaspGlyGlnlyAspAlaGluIleleTyrglyHisAsnA 134  
351 GTACATTATTTATTTGCAAAAAGATGCGGAGATTTATTTATGCGCAATG 400
```

```

134 lathrtlysglgluileuThrThrMetIleGluArgGlyGluThrPasp 150
1401 CACACACAAAGAGAACTACTACATCATATAGACCGAGCAATGGAT 450
151 GluLeuLeuArgArgValValValValProGlyAspPhePheValPr 167
451 GACCTCTGGCGCGTAAAGGTAAGCCGGGATTTTCTATGTCGC 500
167 oSerGlyThrValHisAlaIleGlyLysGlyLeuAlaLeuGluThrG 184
501 AACCGTACTGCTTCATCGCATTCGAAAAGCAATCTGCTTGGAGACGC 550
184 IncInaSenSerAspThrThrThrArgLeuThrAspThrAspArgLysAsp 200
551 AGCAGAACTCAGACACACCTACAGATTATATGATATGACCGAAAGAT 600
201 AlaGluGlyLysLeuArgGluLeuHisLeuLysSerIleGluValI 217
601 GCAGAAAGCAGCTGCGGAGCTTCTATCTGAAAAGACATTTGAAGTAT 650
217 eGluValProSerIleProGluArgHisThrValHisHisGluGlnIleG 234
651 AGAGGTCCGCTTATTCAGAAAGCACTTCAACCTGAACAAATG 700
234 LuAspLeuLeuThrThrThrThrLeuIleGluCysAlaLysPheSerValGly 250
701 AGGATTTGCTTACACGACATGATGATGATGCGCTTACTTTTGGTGGG 750
251 LysThrAsnLeuSerGlySerAlaSerLeuLysGlnIleLysProPhe 267
751 AAATGGAATTTATCAGATCAGACCACTTAAAGCAGCAAAAACCAATCCT 800
267 uLeuIleSerValIleGluGlyLysArgMetIleSerGlyGlyValY 284
801 TCTTATCAGTGTATGAAAGGAGGAGCCGCTATGATCTGCTGACTG 850
284 aLysProPheLysGlyAspHisMetLeuProTyrCylLeuGly 300
851 TCTATCTTCAAAAAGAGATCATATGCTGCTTACGCTTGTGA 900
301 GluPheLysLeuGluGlyTyrAlaGlyCysIleValSerHisLeu 315
901 GAATTTAACTCGAAGATATGAGATGATGCTGCTCCATCTG 945

```

seq_name: gp_baz:BSUB0007

seq_documentation_block:

LOCUS BSUB0007 216750 bp DNA BCT 26-NOV-1997
 DEFINITION Bacillus subtilis complete genome (section 7 of 21); from 1194391
 to 1411140.

ACCESSION 299110 ALD009126

VERSION 299110.1 GI:2633472

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Bacillus subtilis.
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 1 (bases 1 to 216750)
 Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
 Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S.,
 Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
 Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
 Carter, N.M., Choi, S.K., Codani, J.J., Connerion, I.F., Cummings, N.J.,
 Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
 Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
 Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizeli, A.,
 Gallener, N., Gilm, S.Y., Glaser, P., Gottreau, A., Gough, J., Harwood, C.R.,
 Henaut, G., Hilbert, H., Holsappel, S., Hosono, S., Hujo, M.F.,
 Itaya, M., Jones, L., Joris, B., Kobayashi, D., Katsuhara, Y.,
 Kleier, Blanchard, M., Klein, C., Kobayashi, T., Koeltter, P.,
 Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
 Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,

FEATURES
 source
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 The complete sequence of the gram-positive bacterium
 Bacillus subtilis
 Nature 390 (6657): 249-256 (1997)
 98044033
 2 (bases 1 to 216750)
 Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
 Direct Submission
 Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
 Regulation de l'expression Genetique, 28 rue du Docteur Roux, 75724
 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
 68 89 48

Location/Qualifiers
 1. 216750
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 116..1156
 /gene="argC"
 116..1156
 /gene="argC"
 /EC_number="1.2.1.38"
 /function="arginine biosynthesis"
 /codon_start=1
 /transl_table=11
 /product="N-acetylglutamate gamma-semialdehyde
 dehydrogenase"
 /protein_id="CAB12976.1"
 /db_xref="GI:2633473"
 /db_xref="SWISS-PROT:P23715"
 /translation="MKIGIVAGTGYGTELVRLSHHPAEKCIYSSSGEGNYSEG
 YPHLTGLADQOLKPIDMNTIKHEIDIMPLAEPVSSSETPKLAAGATVIDSGDLR
 IKPEAEKWKYKRTAPKAVIOAVYGLAEINQIQOAKLIANGCEPTAVILGDLR
 LAOKRLIDSEFVIDAKTGVSGAKRSGHSELDNRFKIYKNEHOPPEIDRAL
 NEMQDGLPTTFSANHLVPMRTGIMATMYRITCDLADDLIDLYSEFDQSYFVVRP
 KCGYRPRKRYKVSFNCFDIATVLEDRNRTIYSVIDNLKGAAGAVQNFNMGWNE
 ETGLTITRYP"
 1176..2396
 /gene="argD"
 1176..2396
 /gene="argD"
 /EC_number="2.3.1.35"
 /EC_number="2.3.1.1"
 /note="alternate gene name: argA, argE, ornithine
 acetyltransferase"
 /codon_start=1
 /transl_table=11
 /product="amino-acid acetyltransferase"
 /protein_id="CAB12977.1"
 /db_xref="GI:2633474"
 /db_xref="SWISS-PROT:P36843"
 /translation="MTQLSEDOIVKTVGVSSPKGFOANGVHGCLRYSKKDLGVIIS
 TPASAAVYVSHFQAPRIKVDLSLKHGPTLKAVIVNSAINACTGDDGLDAYTR
 ESFASQIGIPELVAVSSTGVIGELDMERIHAGIELLKTPEAGSGDEEALITVDY
 IKQCYELAGKRTVITGARKSGMTHPMATMIGELTDAIDAEKALOKALRETVD
 VSFMOIIVDETNDMVLVMAACENCLIEDHEDMVFYFKKALLLCEIDAKIAR
 DGEATRLIEAVQAGKANNIDANVIYAKIIVGSLVTVAVYGTDAWMGRIGAIHSAA

```

gene
/translation="MKRTYSISILVIGSGPILITIOQAEEPDYSGTQGCIALKEGIVY
LTVYDIKINASYRT"
/EC_number="2.7.2.8"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="N-acetylglutamate 5-phosphotransferase"
/protein_id="CAB12978.1"
/db_xref="GI:2633475"
/db_xref="SWISS-PROT:P36840"
/translation="MKRTYFKCGSVYIRELSEEFYONIKELASGKMLAIYHGSGPE
ITNMLKRNITKTEFSGGQRTKTPVLAEMVYSGSYNKPFFVLAELAKHGLRAAGISGK
DGLLEADYLDPEYGEVGEIKKYDASVMYALMENGIIPIVAPLSMTSDCKTLNVAD
LAASVAGALFADKIMETVYDQIMKEKORLDVLTPEKIQMLIKOEYITGGMIPKYN
ALSALSDQVSEMYVNGKSPFADQTPQGTIKYAKKAVS"
3184.4341
/gene="argD"
/EC_number="2.6.1.11"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="N-acetylornithine aminotransferase"
/protein_id="CAB12979.1"
/db_xref="GI:2633476"
/db_xref="SWISS-PROT:P36839"
/translation="MSLFFQYFGWMDIDIKKAGTYEDQNGKTYLDPIQGLIANSUG
HCHAEVAVKVKOLDVWVHVSNLFONSLQDRAOKLAGHAGDLVPCNNGGAGANCA
YLAARAKGKRTKRTPILOSFGRTYAGMAATGCDKITYGSGPMLAGPHLYPYVDPSPAF
KALGEGCIIAAMVLETVGSEGVNPARSEPIASVOSCKRQXLLIIDEIQTGTGRTG
KRFARHFGSLSDIITVAKGLNGFPGVAYIGKQDLEATPDSHGTTREGNMLAAV
VNAIQLIVQFDPEADKAGFLKEDLELKSPPVKQIRGKGLMIGIJCDDPVADI
IAELDTGLIVIPAGPERDSAAPLVTYKDEIAEVASIKQALAHHSAAVQ"
4412.5473
/gene="carA"
/EC_number="6.3.5.5"
/function="arginine biosynthesis"
/alternate gene name: cpa"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate transferase-arginine
(subunit A)"
/protein_id="CAB12980.1"
/db_xref="GI:2633477"
/db_xref="SWISS-PROT:P36838"
/translation="MEGYLVLEDGTSFSGELDHENCTGRAVFTGNTGYOEVLDPDS
YKGOILVFTYPLIGNYGINEDFESKRPQVYAAVYACDHFHYEAVYSLKLEYLDQK
NIPLTLDVTRAVYKIRANGTGATVATASGEAEIALQENVAEQASQEIETFGDG
NKHILIDFGYKKSIAASLVKGCQKVVYVPOCEAAVYNIKPGDITVSNRGPDKIO
PLIGIKTISIRFPTLIGICLGHOLALAFGNFTKLPFGRRGNHNPIDKTRVMT
SONHSGYVDEOSINEEELTIRFHHVNIOTSVGLAKHKLPMVSNQFHEAHFGAAESEM
ITDDYLVIRIPARRIANA"
5466.8549
/gene="carB"
/EC_number="6.3.5.5"
/function="arginine biosynthesis"
/alternate gene name: cpa"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate transferase-arginine
(subunit B)"
/protein_id="CAB12981.1"
/db_xref="GI:2633478"
/db_xref="SWISS-PROT:P36185"

```

```

/transl_table=11
/product="carbamoyl-phosphate transferase-arginine
(subunit B)"
/protein_id="CAB12981.1"
/db_xref="GI:2633478"
/db_xref="SWISS-PROT:P36185"

/translation="MKRTYSISILVIGSGPILITIOQAEEPDYSGTQGCIALKEGIVY
ILVNSPATIMDEAFADBEIYEPPTAEELIATIKKRPDGLNLNGQTRALNATVLE
EETGVYKREHGVKLLGTSVETIQGDEBREKFRMLKQALNPQESIEVDNEADLHFAE
SIGEPVYIRPAVTLGGKGGIAPDKAEAFNMLKQALNPQESIEVDNEADLHFAE
EVMRDSNNICITVNCNMENIDIPVGHGDSIVVAPQDTLDDYOMLRASLITISALD
VVGSGNIOFADLPFSKQYIVIEVNPVSSSALASKATGYIAKMAALAGYTLDEL
KNPLTGSTVAFEPALDYIVKPEWPFKFNADPKGTMKKATGEVMAIERMLEAA
IKAKASILEKNTGTHPELGSISTDLMDLAITPDDRRPVPVMEILRSYSIDIDHE
KTKIDPELTPFNITIKLENRLMEAGSDSEFLKAKKEDPSATIASISKEEYV
RLRKMEGTPSPKTYDTCAPADPAKTYVYFSGSGPDGDISREKRRALITGSGPI
RIGQGVFEDISAVHGYTLQELGFEYIMNNPVEVSDIYELADLYREPPHTHTYN
VREBENTDFAIVPFGQTAIINAELERKAGTILIGTSOTLDVEDRQYQLDEBGL
LKRAKGEIATVKEEASKASEIGYPLVLPSTYIGKMMITVDSQQLSOLINDESM
PYPLIDQYVSGEVEIDILSDGEVEFIEPTVEHIERAGVSGDSFALLPEPSTSGT
QOQKQVAAOKIARKLSFKIGIMINQVINDGNITVLEVPASRYVPVSKWAPVAPAT
RLAGASLKDNLNPAVONHGVAVKPEVFSNAIQOVDAKLGPEMKSTEGSCVAYDSN
SALKIYTRVWSOKSGIYQNVPEDEYKEIAENAGFTIEGTFASMEQEGNSLHINS
GSEBARERLEAMTHGIPVTEBERVAFIQSGSGHPVSIKIDLYKKEVASCTO"
8537.9510
/gene="argF"
/EC_number="2.1.3.3"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="ornithine carbamoyltransferase"
/protein_id="CAB12982.1"

alignment_scores:
Quality: 1692.00 Length: 315
Ratio: 5.371 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-846-2 x BSUB0007
Align seg 1/1 to: BSUB0007 from: 1 to: 216750

1 MetThrThrGluProLeuPhePheLeuProValPheLeuGluAlaIleTr 17
|||||
79774 ATGACGACATGAAACCGTTATTTTCAAGCCTGTTTCAAGAAATAATGG 79823
|||||
17 pgllygllythrAlaLeuAlaAspPheGlyTyrThrIleProSerGlnIleTr 34
|||||
79824 GGGCGGACCGCCTTTCGCTGATTTGGCTATACGCTTCGTCACAAAGAA 79873
|||||
34 hclglylucystrAlaPheAlaAlaHisGlnAsnGlyGlnSerValIle 50
|||||
79874 CAGGGAGTGTCTGGCTTTTGGCGGCATCAAAATGTCAAAGGCTTGT 79923
|||||
51 GlnAsnGlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGlnIle 67
|||||
79924 CAAACGGAATGATTAAGGGGTTTCAAGCTTCAGCGCATTAATATGGAAATCA 79973
|||||
67 sArgHLSLeuPheGlnGlnLeuGluGlyAspArgPheProLeuLeuTrl 84
|||||
79974 CAACACTTATTTGAGACACTTGAAGGGAGCGCTTCCCTCTGCTTCAA 80023
|||||
84 ylleleuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAsp 100
|||||
80024 AATATATGATGCTGACGACGACTTATCTGTCAAGTGCATCAACAAATGAT 80073
|||||
101 GluTyrAlaAsnIleHisGluAsnGlyLeuLeuGlyLysThrGluCysTr 117
|||||
80074 GAATATGCCAACAATACATGAAAACGTTGAGCTTGAAAACAGAAATGCTG 80123
|||||
117 pTyrIleIleAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsn 134
|||||
80124 GTACATTAATGATTCGCAAAAGATGCCAGATTAATTAATGCGCACATG 80173
|||||
134 lahrhrlysgluGluLeuThrThrMetIleGluArgGlyGluTrpAsp 150
|||||

```



```

RBS
/gene="ydhO"
/feature="SD sequence"
4222..4226
/feature="ydhO"
/feature="SD sequence"
4239..5105
/feature="ydhO"
/feature="similar to B. subtilis ydhO gene(81%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75348.1"
/db_xref="GI:4512385"
/translation="MTAFLSAEYIRTRIVQADPVLGSGIIPPLAIFPIQLMFGVLAG
LTVLAIVFVSQLEHTMMDVITQVADPVLGSGIIPPLAIFPIQLMFGVLAG
OVIINSMVPIWMLTLEMTLEAVTNEPEVPIELFIVMGSGTMTLAVFAI
LAFKSOOLKQVKGIGPGLFVNEPIELFIVMGSGTMTLAVFAI
MASGLVPTGVAVPWVPIFINGIMATNSIAGGLQVLNMLTVMIVMIVPFLKMDRF
NLRQEKESAA"
5104..5111
/feature="SD sequence"
5113..6474
/feature="ydhP"
5113..6474
/feature="ydhP"
/feature="similar to B. subtilis ydhP gene(80%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75349.1"
/db_xref="GI:4512386"
/translation="MPCMTKDTVTVRPEPEFGWGSATANOTEGATHRDGKRNIMH
WVEVPSRFQSGVPODSOPERYKEDIQMKVGNPSFESISMRLLPDGGEVN
KGVFERNVNNIDELGOGIEPPVNLPHDMEMLOKIGMANREVNATVATYKCE
RGGDYKRTFTNRPVYEGGYLDPEYVPEVPEKAIQVGYTVISAKAIRAYO
LNGKRGITLMLTPSYRSNRPADLKAADAFNRSFIDNPAPPEPEPTDTHM
EKGMPITIEEDLEIRQHTVGLGVTIQRRYAKETIDNPAPPEPEPTDTHM
PERKNNVHGWETIEKGIYDILINKETVGNIECISNGVGEKPEFDENVTHD
DYRIEFDHLMKWHRAIQEGSNVKGHLMTFMDMWSANAKNRLWVGQFQENG
TDN"
6505..6509
/feature="SD sequence"
6513..7313
/feature="ydhO"
6513..7313
/feature="ydhO"
/feature="similar to B. subtilis ydhO gene(53%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75350.1"
/db_xref="GI:4512387"
/translation="MDFKRWSYPPDHLPIRINOSLLOKVGCDAMKRYEALISMEIKRIK
ENDYVDPDIPPEISLSKEFGSRMTMKRALDILMEGLYRKRGHGFIVQSIHRS
RVNVLTEKNGKSLNLDKQVTSKYIOENVEFPAEVAHLAIDQKTPVHIIRLV
DEBPVTEIGTVMPTPLIGINDVYHGSYILDKGLTLAGSHRIRACKAKALAI
EHLHCAADPPILEVEHGVFLNNGVPEFSPSRHRRDKREVTIMKKS"
7663..7667
/feature="SD sequence"
7672..8619
/feature="ydhS"
7672..8619
/feature="ydhS"
/feature="SD number"5.3.1.8"
/feature="similar to B. subtilis ydhS gene(53-69% identity)"
/codon_start=1
/transl_table=11
/product="mannose-6 phosphate isomerase"
/protein_id="BAA75351.1"
/db_xref="GI:4512388"
/translation="MLOPTFLPVFKERIMWGTLLKQFQGNISDSKTGECWASIAHP
NNGSIVANGSEIGKSLAELMKQPOLFGNPSDDVPLLTJLIDANMDSLVVHDDST
ANSHEGELGTECWYILDKCKDAEMILGHAKTELEKIEGAGMDLLRVKIKP
GPFYVPGSTIALCEGTIVLETOOSSDTYVYDIDRRDAGNLRHLKKAIDVTT
VPHDDVCPKVTIKDOAVTTFTEFEFSVAKIHQOASFGNSDDRLVLSITEGTG
RLTYGGERXSLKGAHLIPVGFNDFNIEGCOLIVSHP"

```

```

RBS
8061..8063
/feature="ydhS"
/feature="SD sequence"
complement(8719..10722)
/feature="mcpA"
complement(8719..10722)
/feature="mcpA"
/feature="similar to B. subtilis mcpA gene(28%-identity)"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="BAA75352.1"
/db_xref="GI:4512389"
/translation="MISLSEFFKRSILKTVTFPLVGLPIPLIMSTMTQSSNETV
AKERAMSSQVSVSDNEMOELKDEIKIVASSEXLSDRLESKLQIMNTKND
SEIVVFTGAGIVQAHNEHIGVDLSDREYFONGMAGQDTISVLSMTAGNRV
VAVPVLSSQSEITGVMSASINFQILAQLHGDQFQWQHVPIIDQNTIQAHNDEL
IGITVEESGIESDLVSLEATGSEGSTSRVGDSEMLISYAPITERAGYIYLLPLS
SVLATDHIOMITIMAVFTVILVPAIMTASMSRPIRLITSVOTIASGDILSD
LOIKRODEIGILAHNVOMATNLKEVILHOVGDSDORVAAASQOLASSEVSSEOI
SESTIOVASGSEKQSGVNESTLSLEVSSVOOIAASVOITESANSTIEKAKEGGO
SVENNETPMNRTHESVDSVOITKLAERKEININLIDVITSIAEQFNILANATEA
ARAGHGKFAVVDVDEVRKTLAESQSSCOISVIVIOKEMORTSTAMNGVVEEYK
GLTVANETKTRHETIDSTNVABQISMATSEIISAEVQYSASPNELAVTAENTS
DSADQVAAASEQULATMEVSVSAQVYLSQMATDRELKFKV"
complement(10733..10737)

RBS
alignment_scores:
Quality: 988.50 Length: 316
Ratio: 3.831 Gaps: 3
Percent Similarity: 81.646 Percent Identity: 57.595

alignment_block:
US-09-462-846-2 x AB011838 ...

Align seg 1/1 to: AB011838 from: 1 to: 15012

4 GluProLeuPhePheLysProValPheLysGluArgIleTrrpLysGlyTh 20
:::|||||::: |||||:::|||||:::|||||:::|||||:::|||||:::
7678 CAACCGATTGTTTAAACCTGCTTAAAGACCGATTTGGGAGAGGAC 7727

20 FAlaLeu...AlaSpPheGlyTrrHileProSerGlnArgTrrGlyG 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7728 TCTGCTTAAGCAGAGATTGGCTACATATTCATCCGATMAAGCGAG 7777

36 LucYtTrrPalaPheAlaHisGlnArgLysGlnSerValValGlnAan 52
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7778 AGTCTGGCCATTTCGCTCATCCGACAGCGCTCTTATGTTGCTAAT 7827

53 GlyMetTrrLysGlyPheThrLeuSerGluLeuTrrGluHisHisArgH 69
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7828 GCGCAATTATTGGAATAATCATTTAGCAGACTTGGAGAGACGCGCA 7877

69 sLeuPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrLysIle 86
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7878 ACTATTGGAATTCCTCATCAGACGCTTCCGTTATTACGAAAAATTT 7927

86 euAspPalaAspGlnAspLeuSerValGlnValHisProAsnAspGluTrr 102
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7928 TAGACCCCAATATGATTTATCCGTCAGAGTTCAACCCGATGATCATAT 7977

103 AlaAsnIleHisGluAsnGlyGluLeuGlyLysThrGluYstTrrPtyrI 119
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7978 CCGAACACCATGAATAATGGGGAATTAGGAAGACGGAATGCTGTATCAT 8027

119 eLiaspPysGlnLysAspAlaGluIleIleTrrGlyHisAsnAlaThrT 136
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
8028 CCGTATGTTGTAAGAAAGATGCAAGATGATTTATAGTCCACCATGCAAAA 8077

136 hrLysGluGluLeuThrThrMetIleGluArgGlyGluTrrPaspGluLeu 152
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
8078 CGAGAGAAAGAAATTAGCAAAATGATTAAGATGGCGATGCGATGATTTA 8127

```



```

CDS
    /gene="BH3778"
    complement(846..1301)
    /gene="BH3778"
    /note="BH3778
    unknown"
    /codon_start=1
    /protein_id="BAB07497.1"
    /db_xref="GI:10176403"
    /translation="MRKGIACMAVALTCLVGCEREMRPQAOGLNERPAGOVYGLLGP
    GANDNYIKRSIADPAPRONSPSPRAEYPTPTIADIDETISITIDDEHLSRSMY
    MRGSHAVYVVEFPNKKKEORNKIMNDLREDMOKIIRYOVHLSOKG"
    /complement(1382..1966)
    /gene="tok"
    complement(1382..1966)
    /gene="tdk"
    complement(1382..1966)
    /EC_number="2.7.1.21"
    /note="BH3779"
    /codon_start=1
    /product="thymidine kinase"
    /protein_id="BAB07498.1"
    /db_xref="GI:10176404"
    /translation="MOMIKEGMIEYIGSGMSFSGSEELIRVRATYGRKLYQYFKP
    AIDNRYSEGEVSHNGNKVCALPVERAMSLIOWSBDYVAIDVDFEDEVIVVD
    YLADHGVYICAGLDODFRGEPPEPTRLMAEYTKLQALCPVCGSPASRTORLID
    GKPSYHDEIILVGAAYEPRCRHCHOVYTKRSLPETSASSE"
    /complement(2159..2404)
    /gene="rpmE"
    complement(2159..2404)
    /gene="rpmE"
    complement(2159..2404)
    /note="BH3780"
    /codon_start=1
    /transl_table=11
    /product="50S ribosomal protein L31"
    /protein_id="BAB07499.1"
    /db_xref="GI:10176405"
    /translation="MKPGIHPYKVVFPMDSTGFKLSGTSNTSNETIEMEDGNTYP
    LIVERISDSHPFYTKRIADAGRVDRFKKYNLK"
    /complement(2516..3787)
    /gene="rho"
    complement(2516..3787)
    /gene="rho"
    complement(2516..3787)
    /note="BH3781"
    /codon_start=1
    /transl_table=11
    /product="transcriptional terminator"
    /protein_id="BAB07500.1"
    /db_xref="GI:10176406"
    /translation="MGVHLSLENMKRLRELYELARKYKISYYSKITKELIFAILKQ
    AEDGILFMGVLIEIOSGFLRP INYPSSEDIYISASQIRPDLNGKVSQV
    RPKENRYHGLHVEAVNGEDPTKERRHPFALPLPEKRLNETTNGRISSRI
    DMISVPGGGRGLIVAPKAGSILKJANSIVNHYPVELIILLIDRPAENVYDI
    RSYDADVVSSTPEPVENHIVKALVLEBRMRYVEKKKVIILIMOSITPLAAYNVI
    PEGGRILSSGIDPAFAHRRKRPFGAARNTEEGSLIILATAYEGSSBMDVITYEFK
    GTSMELHLDRLSERITFPALDIRSSTRKEELMPKLEKLMVARTMDSPEFY
    DHEIRRYKKTCTDEFEAMEKDMNGTKRK"
    /complement(3913..4257)
    /gene="BH3782"
    complement(3913..4257)
    /gene="BH3782"
    complement(3913..4257)
    /note="BH3782"
    unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB07501.1"
    /db_xref="GI:10176407"
    /translation="MNHRYFEGINEWEAIKITEVEVMEBHCLSIRTPSITIIHGI
    IQYVNNFEFGALSERVIGEIGNHKVQASVLEOAPDRILYLMTRSGFFMCEPSTK
    RTVPDFARMTRD"
    /complement(4164..5129)
    /gene="BH3783"
    complement(4164..5129)

```

```

    /gene="BH3783"
    /codon_start=1
    /transl_table=11
    /product="glycerol-inducible protein"
    /protein_id="BAB07502.1"
    /db_xref="GI:10176408"
    /translation="MERSLSMELVRYTEAALASARWMGNGKNEADRAATEAMRDVF
    DTIPMKGTAVYEGEEMDEAPMLYIGELGNGYPRVAVADPLEGTNIVASGIMNLA
    VIAVADGNLHAPDMYMEKIVAPGVADIDAPYIDLRAVAKKNKVDYLVV
    TINKRDKLIDETROAGARKILIPDDVAAINTAPDDYDGLDLSGSGAPGVLA
    VGKLGELGLOGKLLPOTDEILACKRMAGIEDVNYKVSIDELVKGDCCIPAAVGVDS
    ELKGVYKSTATTOSLVMRAKSGTVRFDPDGKHSKKKRDLYIK"
    /complement(5173..6459)
    /gene="mutZ"
    complement(5173..6459)
    /gene="mutZ"
    complement(5173..6459)
    /EC_number="2.5.1.7"
    /note="BH3784"
    /codon_start=1
    /transl_table=11
    /product="GDP-N-acetylglucosamine
    1-carboxyvinyltransferase"
    /protein_id="BAB07503.1"
    /db_xref="GI:10176409"

```

alignment_scores:

Quality:	988.50	Length:	316
Ratio:	3.831	Gaps:	3
Percent Similarity:	81.646	Percent Identity:	57.595

Align seg 1/1 to reverse of: AP001520 from: 1 to: 305153

US-09-462-846-2 x AP001520/rev ..

```

4  GluProLeuPhePheLysProValPheLysGluArgIleTrpGlyGlyLys 20
   ::::::::::::::::::::|
155351 CAACCGATTTTATTACACCTGTGTTAAAGACGATTTGGGAGGAGC 155302
   ::::::::::::::::::::|
20  fAlaLeu...AlaSPheGlyTyrThrIleProSerGlnArgThrGlyG 36
   |  |||  ::::::::::::::::::::|
155301 TCGCCTTAACGACGAGATTGGCTACATATTCATCCGATAAAGCGGAG 155252
   |  |||  ::::::::::::::::::::|
36  lucystrPalapheAlaAlaHisGlnAsnGlyGlnSerValValGlnAsn 52
   |||  |||  ::::::::::::::::::::|
155251 AGTGTGGGCAATTCGTCTCATCCGACGAGCTCTCTATGTGCTAAT 155202
   |||  ::|  ::::::::::::::::::::|
53  GlyMetLysGlyPheThrLeuSerGluLeuTrpGluHisHisArgHis 69
   |||  ::|  ::::::::::::::::::::|
155201 GCGGAATTTATTGGAAATATCATTAAGACGCTTTGGAAGACGACCGCA 155152
   |||  ::|  ::::::::::::::::::::|
69  sleuPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrLysIle 86
   |||  |||  ::::::::::::::::::::|
155151 ACATATGGAAATCCCTCATCACGCGTTCCTTATTACGAAATTT 155102
   |||  |||  ::::::::::::::::::::|
86  euaSPalaaPglAspLeuSerValGlnValHisProAsnAspGluTrp 102
   |||  |||  ::::::::::::::::::::|
155101 TAGACGCCAATATGATTTAATCCGTGAGGTTCAACCCGATGACTCATAT 155052
   |||  |||  ::::::::::::::::::::|
103  AlAsnIleHisGlnAsnGlyGluLeuGlyLysThrGluCysTrpTyrIle 119
   |||  |||  ::::::::::::::::::::|
155051 GCCAACACGACATGAAATGGGGAATTAAGAAAGACGAAATGCTGTACAT 155002
   |||  |||  ::::::::::::::::::::|
119  eileAspCysGlnLysAspAlaGluIleIleTyrGlnHisAsnAlaThr 136
   |||  |||  ::::::::::::::::::::|
155001 CCTTGATTTGAAGAAAGATCGACAGATGATTAAGTCACCAATGCAAAA 154952
   |||  |||  ::::::::::::::::::::|
136  hrlYsgLugLuleuThrMetIleGluArgGlyLutTrpAspGluLeu 152
   |||  |||  ::::::::::::::::::::|
154951 CGAAGAGAAATTAAGAAATGATTAAGATGCGACATGGATGATTA 154902
   |||  |||  ::::::::::::::::::::|
153  leuArgArgValLysValLysProGlyAspPheTyrValProSerGln 169
   |||  |||  ::::::::::::::::::::|

```

```

154901 TTGGGAAGGTAAGATCAAGCCGCGAGATTTTATGTGCGGAGTGG 154852
169 YTHRVAlHsAlAlleGlySLyIleuAlAluGlUthrInGlnA 186
154851 GAGGATCCAGCACTATGCGAGGGAACGCTCTTAGAACTCAACAA 154802
186 snserApthrThrTYrArgLeuTYrAspTYrAspArgLysAspAlaGlu 202
154801 GCTCGGATACGACGTATCCGGTGTATGACTACAGCCCGAGATGAAGG 154752
203 GLySLyLeuArgGluLeuHsIleuLysSerIleGluValIleGluVa 219
154751 GGAATATTACGGAGCTTCATTTAGAAGGCAAGCCTACGACGACG 154702
219 L.....ProSerIleProGluArgHsIthrValHsIleSLug 232
154701 GCCCATCAAGATCCAGACCTCTGTCGAAAGGACGATAAAGACCAAG 154652
232 InIleGluAspLeuLeuThrThrThrLeuIleGluLysAlaTYrPheSer 248
154651 CCGTT.....GTGACGACGTTGTAGAAAGGAGTCTTTTCC 154614
249 ValGlyLeuThrPAsnLeuSerGlySerAlaSerLeuLysGlnGlnIlyPr 265
154613 GTCCATTAATGGACATTCACGAGAGGCTTCCTCTCTAGTGACATCG 154564
265 oPheLeuLeuIleSerValIleGluGluGluArgMetIleSerGlyG 282
154563 TTATTTACTGTAGATCATGTGAGAGACGAGCTTACTATCAAG 154514
282 LuTYrValTYrProPheLysGlyAspHsMetLeuLeuProTYrGly 298
154513 GCGAGACATCTCTTAGAAAAGGCGCCACCTCTAATCCAGTAGAGG 154464
299 IeugIlyLupheLysLeuGluGlyTYrAlaGluCysIleValSerIls 314
154463 TTTAAATTTTAACTGAAAGCCCAATGCCAATTGATGCTTCCAT 154416
seq_name: gb_pat1:AX006203

seq_documentation_block:
LOCUS AX006203 948 bp DNA PAT 24-AUG-2000
DEFINITION Sequence 4 from Patent WO9904016.
ACCESSION AX006203
VERSION AX006203.1 GI:9929072
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis.
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE
1 (bases 1 to 948)
AUTHORS Estell,D.A.
TITLE Proteases from gram-positive organisms
JOURNAL Patent: WO 9904016-A 4 28-JAN-1999;
GENENOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)
FEATURES
source 1..948
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 286 a 192 c 244 g 226 t
ORIGIN

```

```

alignment_scores:
Quality: 978.50 Length: 316
Ratio: 3.822 Gaps: 1
Percent Similarity: 81.013 Percent Identity: 56.646

```

```

alignment_block:
US-09-462-846-2 x AX006203

```

```

Align seg 1/1 to: AX006203 from: 1 to: 948

```

```

1 MethTrhrGluPheLeuPheLysProValPheLysGluArgIleTr 17
1 ATGAGCGCATCCCGATTCTTACCGCTGTGTAAACAAAATCTG 50
17 pGlyGlyThrAlaLeuAlaSP...PheGlyTYrThrIleProSerGlnA 33
51 GGGCGGAACCCCTTACGAGATAGATTGGATGACAGATTCTTCAGAA 100
33 rYrThrGlyLysTrpAlaPheAlaIleHsGlnAsnGlyIleSerVal 49
101 CAAGCGGAGATGCTGGCCATTCCCTCATCCAAAGAACCGACACT 150
50 ValGlnAsnGlyMetTYrLysGlyPheThrLeuSerGluLeuTrpGluH 66
151 GTTCAATAGCCCGTATTAAGAAAGACATGTATGACGTTTGGGAAGA 200
66 sHsArgHsIleuPheGlySLyLeuGluGlyAspArgPheProLeuLeu 83
201 GCACCGTAAGTATTCGGCGGTAGAGGGGATCGGTTCCGCTCTGA 250
83 hTrLysIleLeuAspAlaSPGlnAspLeuSerValGlnValHsProAsn 99
251 CAAGCTGCTGATGTAGAGAAATACGTCATTAAGTTCACTCAT 300
100 AspGluTYrAlaAsnIleHsGlnAsnGlyLysGluGlySLyThrGly 116
301 GATTACTATGCGCGGAGAAAGCAAGAGGAGAACTCGCAAGAGGAATG 350
116 sTrpTYrIleIleAspCysGlnLysAspAlaGluIleIleTYrGlyHsA 133
351 CTGTACTATTCACGTGTAAAGAAAGCAAGAAATCATTTAGCGCAT 400
133 snAlaThrThrLysGluGluLeuThrThrMetIleGluValGlyGluTrp 149
401 CGGCCCGCTCAAAAACGAACTTCTCAGATGATCAACACGGGTACTG 450
150 AspGluLeuLeuArgValLysValLysProGlyAspPheThrTYrVal 166
451 GAGGCTGCTGCGAAGAAATCAAAATTAACCGGTCATTTCTATTATG 500
166 lProSerGlyThrValHsAlaIleGlySLyGlyIleLeuAlaLeuGlu 183
501 GCCAGCGGAAGCGTGCAGCATGTGCAAGGGGCGCTGTATTAGAGA 550
183 hTrGlnLysAsnSerAspThrThrTYrArgLeuTYrAspTYrAspArgLys 199
551 CTCAGCAAAATTCAGATGCCACATACCGGGTGTATGACCGTCT 600
200 AspAlaGluGlyLysLeuArgGluLeuHsIleuLysSerIleGluVa 216
601 GATGACAAAGGAAAGTCCGAGAGCTTCATTGTGCCAAAGCGGCAATGC 650
216 lIleGluValProSerIleProGluArgHsIthrValHsIleGluGln 233
651 CGCCAGGCTTCCCATGTGAGCGGTATATAGATGATCAACAGAAATCA 700
233 lGluAspLeuLeuThrThrThrLeuIleGluCysAlaTYrPheSerVal 249
701 GAAAGAAATATACATTAAACATTTGTCCAGGGAATATTTTGGCT 750
250 GlySLyTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnIlySProph 266
751 TATAAATGGACATCAATGGGAGAGCTTGAATAGCTTCAGATTAATCT 800
266 eLeuLeuIleSerValIleGluGlyGluArgMetIleSerGlyLys 283
801 TCTGATTTCAGCGCTGATAGAAAGCAACGCTTGTCAATATAGACACA 850
283 YrValTYrProPheLysSLySLyAspHsMetLeuLeuProTYrGlyLeu 299
851 AAACATGTCGCTCAAAAAGGTATCATTTATTTGCCGGGTCAATG 900
300 GlyLupheLysLeuGluGlyTYrAlaGluCysIleValSerHsIleu 315

```



```

636 GAGGCGCTGCGGAGACATCAAAATTAAACCGGTGATTTCATATATGT 685
166 1PserGlyThrValHisAlaIleGlySerGlyIleuAlaIleuGluT 183
686 GCGGAGCGGAGAACCTGCACCATTTGTCAGAGGGGCCCTGTTTATGAGA 735
183 hrgIngaInaenSerAspThrThrTyrArgLeuTyrAspIlyAspArgLys 199
736 CTCGCAAAATTCAGATGACCATACCGGCTGATGACATTTATGACCTGTT 785
200 AspaIaGluGlyLysLeuArgGluLeuHisLeuLysSerIleGluVa 216
786 GATAGCAACGCAAGCTCCAGAGAGCTTCATTTGCCAAGCGGTCAATGC 835
216 111G11Val1ProSerIleProGluuArgHisThrValHisHisGluGlnI 233
836 CGCCACGCTTCCCATGTGACGCGGTATATAGATGAAATCGACAGATCA 885
233 legIuaSplLeuLeuThrThrThrLeuIleGluGlyAlaTyrPheSerVal 249
886 GAAAGCAATTAACCATTAACATTTGTCAGAGGCAATATTTTTCGTT 935
250 G1lyStrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPh 266
936 TATAAATGGGACATCAATGCGAAGTGAATGCTCAGATGAATCTT 985
266 eleuLeuIleSerValIleGluGlyLysGlyArgMetIleSerGlyLut 283
986 TCTATTTGCGCCTGATAGAGAGCGGCTTGTCAATATAGAGACA 1035
283 yVal1TyrProPheLysGlyAspHisMetLeuLeuProTyrGlyLeu 299
1036 AAACATGTCGCGTCAAAAAGGTGATCATTTATTTGCGCGCTCAAAAG 1085
300 G1yGluPheLysLeuGluGlyTyrAlaGluGlyIleValSerIleu 315
1086 CCCCATTTACGATAAAGAACTGTACCTTATGCTGTCATCATTT 1133

seq_name: gb_ba2:BACORFX

seq_documentation_block: 4162 bp DNA BCT 01-FEB-2000
LOCUS BACORFX Bacillus subtilis gene for beta-N-acetylglucosaminidase, complete
DEFINITION cds.
ACCESSION D45048.1 GI:1129072
VERSION D45048.1 GI:1129072
KEYWORDS endo-beta-1,4-glucosaminidase; beta-N-acetylglucosaminidase.
SOURCE Bacillus subtilis (strain AC327) DNA, clones PEH4 and pMCEV4.
ORGANISM Bacillus subtilis
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Bacillus.
1 (bases 1 to 4162)
REFERENCE 1 Margol.P., Manuel.C. and Karamata.D.
AUTHORS The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168
TITLE Cell. Microbiol. 12 (4), 535-545 (1994)
JOURNAL 2 (sites)
MEDLINE 95020588
AUTHORS Rashid.M.H., Mori.M. and Sekiguchi.J.
TITLE Glucosaminidase of Bacillus subtilis: cloning, regulation, primary
JOURNAL structure and biochemical characterization
MEDLINE Microbiology 141 (Pt 10), 2391-2404 (1995)
REFERENCE 3 (bases 1 to 4162)
AUTHORS Sekiguchi.J.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1995) to the DDBJ/EMBL/GenBank databases.
MEDLINE Department of Applied Biology/ 3-15-1 Tokida, Ueda-shi, Nagano
AUTHORS Japan (E-mail:jsekiguchi@ipc.shinshu-u.ac.jp, Tel:0268-21-5344,
FEATURES Fax:0268-21-5331)
source 1. 4162 Location/Qualifiers

```

```

-35_signal
-10_signal
RBS
gene
CDS
stem_loop
-35_signal
-35_signal
-10_signal
-10_signal
-10_signal
RBS
sig_peptide
gene
CDS
mat_peptide
/organism="Bacillus subtilis"
/strain="AC327"
/db_xref="taxon:1423"
/clone="PEH4 and pMCEV4"
34...39
/note="Sigma-A type"
61...66
/note="Sigma-A type"
200...208
/gene="orfX"
217...1167
/gene="orfX"
217...1167
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BA08088.1"
/transl_table=11
/db_xref="GI:1129073"
/translation="MTQSPITLPVPEKRTIMGCTALRDPRGYSIPSESTGECMAISAH
PKGSEYVANGYKKGRTILPMERREYGVGVEDRDLPLKLLDVEEDPSIVHPDY
VAGNERBDELKTCWYIIDCKENAEIYHTARSTELIYMTNMGDMGILIRIKY
PGRIVYPSGLHALCKGALYLEIQDSDAITVYIDYDLDSGSRLEHFAKAYNA
IVPHVDGIDESTSKRGITIKTFVQGEIVSVIRWMDINGEAKADESFLICSVLDGS
GLKYEDTCEPLKRGDHPILPQMPDFTIGTCTLVSR"
1166...1194
/note="rho-independent terminator"
/evidence="not_experimental"
1224...1228
/note="sigma-D type"
1241...1246
/note="sigma-A type"
1244...1250
/note="sigma-D type"
1265...1270
/note="sigma-A type"
1285...1291
1296...1376
/gene="cw1G"
/product="27 amino acids long signal peptide"
1296...3938
/gene="cw1G"
1296...3938
/gene="cw1G"
/EC_number="3.2.1.30"
/function="hydrolyze beta-1,4 glycan bond between the
GlcNAc & MurNAc"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="beta-N-acetylglucosaminidase"
/protein_id="BA08089.1"
/db_xref="GI:1129074"
/translation="MKRRILNPMILSAASLAFPMGSGNQAAYTDYSIKYKPSNFT
STESGASQAVAKLEKDTGMDASQYDSITTTIQTQASGHISEBAKAIISGLAKQSTI
TDIISPVSQKPYLTISGATSGEKQNTILAKIKOETGAVAGVAKIAGAPPMWMT
SGIADERTVKALIQSLAKQGTGKSSYQPIHTYVSVTTIGSTGVGSRQAQIENAFQK
ESGLOASLEKTEVKGQAVYETTAIISBANAKTLLDQLKQSTGITSYSINQKTVE
SYNVOASLEKSTVKDAISQIKKNTVSGSYQVQVCSYVNMKGITRQDQKIDT
FEKKKWHYVSSVSKTTSNAVOITTAKLIGEQANKAAFPQKRVAAKTAAGST
AENQQLISEETSOQAVTKQILNKLKQLSASAKSVKQIADTFKRTTESLIDQKY
NOALTEFSNHSIVASOKTGOTASQYQITTEALISQEEIDRVLTFFKQNHVADTSK
TGOATVQYKIVTTOLSKTRALNGLYLSKSVTPETYSNTLTKISVNEDEFTND
TAAASRTKLQLYGNTSSIVKIKNGPQIMKTNTNLSLRDVQKQMTVSPOTGAAYVS
LTYINTASTYATAVINIRSPREYSPNPGVQKRGKISYVQELKISYEVNKGVLAKGILTGPAKFI
SDRYVOYVDPNNSRDSKRYFQELKISYVQELKISYEVNKGVLAKGILTGPAKFI
AANQYSIEHLIISHALLETNGTSLANKLITNGKTYVMKTYIGAYDSNPYQYK
ATQGWTFPEKAIIGAKFTQSSYIHTATVATNODILYKRMASATATQVATIDGMWIKQ
VNRKISLSDLDGTLTFDPEYR"
1377...3933
/gene="cw1G"
/EC_number="3.2.1.30"
/note="853 amino acids long mature peptide"
/function="hydrolyzes beta-1,4 glycan bond between GlcNAc

```


TITLE Direct Submission
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES Location/Qualifiers
1. 212610
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(10. .3879)
/gene="yvnB"
complement(10. .3879)
/gene="yvnB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15521.1"
/db_xref="GI:2636030"
/db_xref="SPTREMBL:O34986"

gene
CDS
/translation="MRKYTVASILLISPLSYLSGGHESKAFVVOGELRPPHYIPL
LENPOIDVKAEMSNLMEAOVRDINSYOAFAFLFKOSNELCYKRVMPDPGAVR
KFIQIPKRIITNSELRYITMSNGKORSEETKILLEGQADLAHIFPELLTELAV
DTKNIGRADGTEFIEFTTTRTDYDFKQYHRYRYPRKEGOSDLIMHEDRILTSGE
TFVVALPAGHELEISADPNRTQTQLEKGNLAVIDETROMANTPRRAVISTNTEK
DISVAHTRKHLRLSSVLRKPLNGLAELNLSIGERNPSGTVLAQVDPDKRIK
LDREKPVLEDLDRKVPRAESIELRADIDRSISVKTAVAFYRTDENKPKRILAEKD
RNDLSEHTIVYSPELIGKQLEYVAAVDG:INEARTKMDIKQTSKAGLNMNEN
RDTLSGTQFLATTEGRADS:IKLMDIKQVTEPAKMEYVFAFDKRIKLYKRNVT
MEGKLVAFDDTNTKRYTSVLPETILRKQKQIQRITIRSGSVSPEDFAENRDEL
KGNALVSDGTVIDRQVSEKELFTGDNQSNKSMQFQNLDPGLFTQLEMDIS
KLESGAHIIQASDGENSVLVYVNSGPHIEPNITGQYKGNLILQADAMYKWSIT
EEASGLDSEITLPLYHTSSDLLPGKHSLEKVTADLAGNKTVIERLEKTERHPDP
EVIDEADNTHAKLSVRYKQDPTNDAMPDVGFRGQYARHNVKIFKIASLDEPKSY
PEREPTPTNKELEYSADGKSTENKELFPRYARFVTVDPSSIDENDLAEVYKSS
LPGRKVTMYANNRYTNEMQPYDSFAKDDKPFITKASVIADAPVRESKNAVYODEP
PAKMDYFVWMSDPOYAESEYPTDEKQTEKIRNKOINIKYVHTGDIYDSDSADP
OKNADBSMSVLDGTPYGLAENHNGHSDSYRAFGRGSDRDKDFHGGSK
NNRGHYDITSSNGNDYIMLNGWITBEDLMMNQVLRKHDRMALTAEHYELVSGN
NSPIGEKLEKIVKPNPVAVLSGHTSHAKRTLEDDDGDKPDLVHOMLADYOG
GVEGGQYLRLLQFDQANDVAVSTSPYVDRKMYVDTDYGNKDEFSISLDIKPRK
KVEDYFECNVYTNELGRKQVSGTAERFMDLEPQSVYVYIIVEDSFGSKRS
PIWFKETKREYRPADQDFDRHVSNP"

gene
CDS
/gene="yvnA"
4044. .4517
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB15522.1"
/db_xref="GI:2636031"
/db_xref="SPTREMBL:O34692"
/translation="MDSIHDELFOALQIOFALKRDKRVOKVOPTSGISSQHNDHK
KDWYTLQAHVSCIHSTSONVNSFLASRLHISKAVASKAVALKHNIIVTKRPGK
KEIFTYLTDSGRKIALHEQJHEKAKQYKOLFNEFSIDDLKVTATFENIMIKYM"

gene
CDS
/gene="cyuX"
complement(4558. .5775)
/gene="cyuX"
/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="CAB15523.1"
/db_xref="GI:2636032"
/db_xref="SPTREMBL:O34926"
/translation="MSOSIKLFSVLSQFQNNPYAFPSQLEEDPVHYEESIDSYIS
RYHDVRYILOHPDIFPTFSKLEPVRGCVLAQMGRKSHAKRRIVVFSFGALD
HLSPLIKONANENLAPYLERGKSDLVNDFGTFVAVCYTMDMLGDKDRKIKSEMG
VADFTTISQSPERASIMSEQLSYIMPYIKERVNPNGSDLSITLCSEVEGAL
SDKDLALTLNVLAALEPDKTLAMIMHILNNPQDMNDVLDASLVPRAIMETLRY

gene
CDS
/translation="KPPVOLIIPROLSODTVGMEIKKDTIVCEMIGANRDPPEAFDPVFNIRREDLGIK
SAFSGAARHLAFSGSIGHNCVGAFAKNEIEIVANIVLDKMRNIRLEDFCYAESGLYT
RCPVSLIAVAFDA"
complement(5791. .6537)
/gene="yvmC"
complement(5791. .6537)
/gene="yvmC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15524.1"
/db_xref="GI:2636033"
/db_xref="SPTREMBL:O34351"
/translation="MTGMVTERRSYHFAEALTFENCREIEERRRHLVIGISPNRPS
EDYIYRLIKMAAQFQSVSLAGHEANLLEALGTPRGKAERKVRKESRNRFAER
ALVAGDPRKAIHTFSDPIDNKAYQLRQEVAFEPQPHRAKCLDSREALIGRAR
GVSLMEVSEEDMLNLAVEYVIAELPFIPGADILEVEETLLAVHRPMKGEKISNHE
FSICMRPNQGYIVQEMQMLSEKRIISG"

gene
CDS
/gene="yvmB"
6963. .7472
/function="unknown"
/note="alternate gene name: yzha"
similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15525.1"
/db_xref="GI:2636034"
/db_xref="SWISS-PROT:P40762"
/translation="MSDLTKOMIYDIYVRLHLNEOKANSLIOOFFKEAEEDVAEIP
KMTSIIHAYDCGHEFIRNNIGAKRMLSKANYTKISTRLKEEFINSYOLDNKE
VYFKLTRGGRICFDLHEHLKKEELAFYQGLFDSQGEQKAVLKFLQLNSTLEAGET
DGTPTPKPK"

gene
CDS
/gene="yvmA"
7493. .8726
/gene="yvmA"
7493. .8704
/gene="yvmA"
/function="unknown"
/note="similar to transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB15526.1"
/db_xref="GI:2636035"
/db_xref="SPTREMBL:O34307"
/translation="MAHTKSKAVLILTVCSAFASISONIVSPLIPIIKESPHVST
AMVNLSVSEVMVTAIMOITIGATIDREGARIVITGITLNRASISGCAVTDPTLTL
IFRMIOAGSALPILIAATTTGOLFTGERSAMGTOMLISVAPALAPVIGFIGA
ACYEGIFWILASISVLYVNSITFPKDSPTESQOKGNVFAHKSIFINRIGNVYL
ILSPVETIFRAVITYPLILTEHRTIDVAGLLYPLASTIAGTFLKRIQAKIG
LHTPLFGSNVIAACSIILFAVTHSVSLVLMALTLALFGLISGVTIPPLYSTWITNEEH
NRGSAIGMNFIRYTGMAAPVSAVLLTTPMSMPSLLGLGFAALSFCLLPMSSP
KRTKGRKHHM"

terminator

gene
CDS
/gene="yvmA"
8709. .8726
complement(8732. .9091)
/gene="yviD"
complement(8732. .9091)
/gene="yviD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15527.1"
/db_xref="GI:2636036"
/db_xref="SPTREMBL:O34648"
/translation="YKNAVASILVNALLVIVIDYIDSIRHISGAIINSLITLILN
VHLIKLITFLPEVIMVYLGFLFVINAITLMTASIMGDSFOIDGCGTAIMASVILLS
VFHLIKIGLIEPLRRK"

gene
CDS
complement(9093. .9290)
/gene="yviC"
complement(9093. .9290)
/gene="yviC"
complement(9093. .9290)
/gene="yviC"

```

/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15528.1"
/db_xref="GI:2636037"
/db_xref="SPTRMBL:034719"
/translation="MNKLYRSEKNKTIAGVIGLAEXFNMWDSLRLVITVILATMTSV

```

alignment_scores:

```

Quality: 978.50      Length: 316
Ratio: 3.822         Gaps: 1
Percent Similarity: 81.013      Percent Identity: 56.646

```

alignment_block:

```

US-09-462-846-2 x BSUB0019/rev ..

```

```

Align seg 1/1 to reverse of: BSUB0019 from: 1 to: 212610

```

```

1 MetThrThrGluProLeuPhePhePhePheProValPheGluArgIleT 17
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
90517 ATGACGCAATCACCGAATTTTCTACGCCCTGTGTTAAAGAAAATCTG 90468

17 pgiLygLyThAlaLeuAlaAsp...PheGlyTyThrIleProSerGlnA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90467 GGGCGGGAACCGCTTACGAGATGATTGGATACAGATTCTTCAGAAAT 90418

33 rgrThrGlyGlyCysTrpAlaPheAlaAlaHisGlnGlnSerVal 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90417 CAAGGGGGGAATGCTGGCCATTCTCCGTCATCCAAAGAGCCGAGACT 90368

50 ValGlnAspGlyMetTyThrGlyPheThrLeuSerGluLeuTrpGluH 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90367 GTTGCAAAATGCCCGGTATTAAGAAAACACTTGTATCGACCTTGGGAGA 90318

66 SHSArgHisLeuPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90317 GCACCGGGAATGTCGCGCGGTAGAGGGGATCGGTTTCCCTCTGTA 90268

83 hTrpSileuAspAlaAspGlnAspLeuSerValGlnValHisProAsn 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90267 CAAGCGCGCGGATGTGAAGAGATACGTCATTAAGTTTACCCCTGAT 90218

100 AspGlyTyThrAlaAsnIleHisGlnGlyGlyGlyGlyGlyGlyGly 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90217 GATTACTATGCGGGAAGAAACGAGAGGAGAACTCGCAAGACGGATG 90168

116 strPTyrlIleAspCysGlnGlyAspAlaGluIleIleTyroGlyHis 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90167 CTGGTACATATATCGATGTAAAGAAACGAGAAATCATTTACGGCATA 90118

133 snAlaThrThrLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90117 CGGCGCGCTCAAAAACGAACTTCTCAATGATCAACACGCGTACCTG 90068

150 AspGluLeuLeuArgValValValValValValValValValValVal 166
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
90067 GAGGCGCTGCTGCGAGAAATCAAAATTAACCGGCTGATTTCTATTTGT 90018

166 lProSerGlyThrValHisAlaIleGlyGlyGlyGlyGlyGlyGlyGly 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90017 GCGGAGGGGAACGCTGCGACGATGTGCAAGGGGCGCTGTTTACAGA 89968

183 hTrpGlnHisAsnSerAspThrThrTyThrGlyGlyGlyGlyGlyGly 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
89967 CTCGACAAATTCAGATGCCACATACCGGGGTGATGATTAAGACCGCTT 89918

200 AspAlaGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
89917 GATAGCAACGGAAGCTCCGAGACCTCATTTTCCAAAGCGGTCAATGC 89868

216 lIleGluValProSerIleProGluArgHisThrValHisHisGluIn 233
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
89867 CGCCAGGTTCCTCCATGTGACGGGTATATGATGAATCGACAAATCA 89818

```

```

seq_name: gb_pat1:AX006205

```

```

seq_documentation_block:
LOCUS AX006205 945 bp DNA PAT 24-AUG-2000
DEFINITION Sequence 6 from Patent WO9904016.
ACCESSION AX006205
VERSION AX006205.1 GI:9929073
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
1 (bases 1 to 945)
AUTHORS
Estell,D.A.
TITLES
Proteases from gram-positive organisms
JOURNAL
Patent: WO 9904016-A 6 28-JAN-1999;
GENEOR
INTERNATIONAL B V (NL); ESTELL DAVID A (US)
FEATURES
location/Qualifiers
source 1..945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 290 a 189 c 223 g 243 t
ORIGIN

```

```

alignment_scores:
Quality: 958.50      Length: 313
Ratio: 3.804         Gaps: 1
Percent Similarity: 80.511      Percent Identity: 55.591

```

```

alignment_block:
US-09-462-846-2 x AX006205 ..

```

```

Align seg 1/1 to: AX006205 from: 1 to: 945

```

```

3 ThrGluProLeuPhePhePhePheProValPheGluArgIleT 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||
4 ACCATCCATATATTTTAGAGCCCTGCTTAAAGAAAGCTATGAGGAGG 53

19 yTrpAlaLeuAlaAsp...PheGlyTyThrIleProSerGlnArgThr 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||
54 GACGAGCTTCGTCGACGCTTTGGTACGCAATACCTCCAAACAAACAG 103

35 lYglGlyTrpAlaPheAlaAlaHisGlnAsnGlyGlnSerValValGln 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||
104 GTAGTGTCTGGCGCTTCTGCAATGCCATGCTGCTGCTGTGTAAGA 153

52 AsnGlyMetTyThrGlyPheThrLeuSerGluLeuTrpGluHisHis 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||
154 AATGCGCGCGTGGCAGGAAGACACTTGTATGATCAAGTATGCAATCAT 203

68 gHisLeuPheGlyGlnLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 85

```



```

85  leleuaspAlaaspGlnAspLeuSerValGlnValHisProAsnAspGlu 101
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18079  TGGCGAGCGCCAAATGATGATCTCTCCGTGCAATGATCTCTGATGATGAT 18128
102  TyrAlaasnIleHisGluAangGlyGluLeuGlyLysThrGluCysTrpTy 118
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18129  TATGCAAAACCTGCACGAAATGGCAGCTTGTAATAACGAGAGTGTGGTA 18178
118  rIleIleAspCysGlnLysAspAlaGlnIleIleTyGlyHisAsnAlaT 135
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18179  TATCATTCATTCGCAACAGATGACCGCGAATATTTGGGACATCATGCA 18228
135  hrThrLysGluGluLeuThrThMetIleGluArgGlyLutPaspGln 151
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18229  GCACAAAGAGAGAGTTCAACCAACGATACAAAGCGTGTGTGGAACGGC 18278
152  LeuLeuArgArgValLysValLysProGlyAspPhePheTyValProse 168
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18279  CTGGTGAGCGCAATCAAAATCAAGCCGAGGATTTCTTTATGTGCGCAG 18328
168  rGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGln 185
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18329  CGGACACCTCATGCTTATGCTTAAGGACCCCTTGCTTGTAATCCAG 18378
185  LmsAsnSerAspThrThrThrArgLeuTyAspTyAspArgLysAspAla 201
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18379  AAAACCTGTATACAAACATATCGGCTATACGATATATGACCGCTGTAATGAC 18428
202  GlnGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleG 218
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18429  CAGGCGCAAAAGAACTCTTCATATGAAAAGCATGGAATCATATAC 18478
218  vAlaProSerIleProGluArgHisThrValHisHisGluGlnIleGluA 235
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18479  GATACCCCATTCGATTAAGTCACTACACCGGAAGTAAGAAGATGGTA 18528
235  sPheLeuLeuThrThrLeuIleGluCysAlaLysPheSerValGlyLys 251
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18529  ACGGTGATCATGTTGTTATGTGCATACAGATATATTCGACGTGTAACA 18578
252  TrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeu 268
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18579  TCGAAGATTTAGCGCGCGACGCTCTTCCTCATATCAAAACCATATTC 18628
268  uIleSerValIleGluGlyLysGluArgMetLieserGlyLysValT 285
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18629  GGGGAGTCTCTGACGCGATCAGACGAATCAATTAATATGATTCAGT 18678
285  yTrpProPheLysGlyAspHisMetLeuLeuProTyArgLeuGlyGlu 301
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18679  ATGAAATCAATGCAAGGCTCACACTTATATCTGCTCGCCATTTTGGAGA 18728
302  PheLysLeuGluGlyLysValGluCysIleValSerHis 314
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18729  TTTTACATAGAGGAACATGTGAATTCATGATATCTCAT 18767

seq_name: gb_baz:BSUB0004
seq_documentation_block:
LOCUS      BSUB0004      213190 bp      DNA      BCF      26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 4 of 21): from 600701 to
            813890
ACCESSION  Z99107 AL009126
VERSION    Z99107.1 GI:2632866
KEYWORDS
SOURCE
ORGANISM   Bacillus subtilis.
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus subtilis
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  1 (bases 1 to 213190)
            Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
            Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,

```

```

Boriss, R., Boursier, L., Brans, A., Braun, M., Bignell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehlich, S.D.,
Emmerson, P.T., Enlian, K.D., Errington, J., Fabret, C., Ferrati, E.,
Fouger, D., Fittz, C., Fujita, M., Fujita, Y., Funa, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Goldlighty, E.J.,
Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kaishata, P.,
Klaerr, Blanchard, M., Klein, C., Kobayashi, T., Koetter, P.,
Kontingstein, G., Krogh, S., Kumano, M., Kurita, K., Lapius, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, D., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, D., Portwillik, S., Prescott, A.M., Prescan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieser, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serron, P., Shin, B.S., Solido, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Yamakoshi, A., Yanaka, T., Terpsstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandendol, M., Vannier, F., Vassaret, A.,
Viart, A., Wampert, R., Wedler, E., Wedler, H., Weltzenegger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.

The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
2 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
            source
            Location/Qualifiers
                1..213190
                /organism="Bacillus subtilis"
                /db_xref="taxon:1423"
                33..740
                /gene="ydf5"
                33..740
                /gene="ydf5"
                /function="unknown"
                /note="similar to hypothetical proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12373.1"
                /db_xref="GI:2632867"
                /db_xref="SPTREMBL:P96697"
                /translation="MIELEVIRIVASFGILLIERIKOTISOMTIFPIAATIG
                ATRAGAKVTSIKPHNMAISFISIVLITFLIFLISIKRKRFKPADPVLIONGI
                LESNRKMYRLDYLNQOLREKEIPEVEFATLENTGOLTVLRKQFHYVKODIM
                IAVNQRPIELMDGELIENLNKONLIESWLELRKRDIVKXETVAVLGNQD
                IYVDYKHISYPMKR"
                complement(755..1186)
                /gene="ydfT"
                complement(755..1186)
                /gene="ydfT"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12374.1"
                /db_xref="GI:2632868"
                /db_xref="SPTREMBL:P96698"
                /translation="MDPEKIRKWEITITRYOSDFWTVLYKAPKPEFPOSEASTPY
                DTQDEINFTIVEMPVIEELITRLSKTQLLIKITITPFAEMSEVLERIKETIGEL

```

gene
complement(1199. .1441)
/gene="ydgq"
complement(1199. .1441)
/gene="ydgq"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=1
/protein_id="CAB12375.1"
/db_xref="GI:2632869"
/translation="MPSTVINYLYLKINSISGNSITIGEAAYNSPTNNOKSOGTSS
GDLPTSSILNTGNDSDVDODIGNSAPISNQI"
complement(1455. .1727)
/gene="ydgq"
complement(1455. .1727)
/gene="ydgq"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=1
/protein_id="CAB12376.1"
/db_xref="GI:2632870"
/translation="MPSTVINYLYLKINSISGNSITIGEAAYNSPTNNOKSOGTSS
FGDTSPTSESVENFLNDPDVNDQTSIGNSDTSVNPAPIAPPILD"
2026. .2613
/gene="ydgq"
2026. .2613
/gene="ydgq"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=1
/protein_id="CAB12377.1"
/db_xref="GI:2632871"
/translation="MKSQGVSRKRLKAAANFVSRGFHDAKVEIYKAGTQPSF
LYLQSKAIPAEILITDFHSVRKLESLLENGLTVEVSKVLLAVETVQFDE
KDLKRVGFLEPBAKQMKDLAVYLKENLEAQRGLYFSEIDMETVACLVGMTEHL
TESFLITGIDKPAISLADEVNLLIYGLPFGNDVR"
join(2610. .2954,3074. .3091)
/gene="ydgq"
2610. .2954
/gene="ydgq"
/function="unknown"
/codon_start=1
/transl_table=1
/protein_id="CAB12378.1"
/db_xref="GI:2632872"
/translation="MISIMKVSILAVFMLAGGIKYSRVPEVHEHMYOYELMLTV
TGILEAGALMTAGIMNRYAIGAGVLFVVLNAGAIHMHFRAROSVIMAIOMICL
IVSIIMIMGSTY"
3074. .3091
/gene="ydgq"
3117. .3590
/gene="ydgq"
3117. .3590
/gene="ydgq"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=1
/protein_id="CAB12379.1"
/db_xref="GI:2632873"
/translation="MAFTLEDMTEEEFEAFRGMSYONFAKONTTSCITWTEKPAEKS
OAYENMLDNGSSNHFWNTINDOGERMGLVADPHLPQKAFVSGVGLYEA
KGLAQLQTLDERARRELGARLALHVFANETAVYLXOMGYAMTNIMRQLC"
complement(3705. .3722)
/gene="exp2"
terminator

gene
complement(3705. .5393)
/gene="exp2"
complement(3750. .5393)
/gene="exp2"
/codon_start=1
/transl_table=1
/product="ATP-binding transport protein"
/protein_id="CAB12380.1"
/db_xref="GI:2632874"
/translation="MKEIVLTNYSVEYKQVTEKRVNNSVQOGDITIGIKNGAGKS
TLIHLINDLAPAQQLIKDLALVEQETAAVSFADQTPAKKILKEMHVPDLDFH
QLSGEKKARLAKGLSEADLLIDEPYHLDKSLQFLQOLKHYNTVLLVSDR
YFLDEARVINSLEQTLIEFGKNSGVGMFKREKRILTOREYKQKAVRELOMN
GLASWEKRAHAGSTKEGFEYHVRKAKRTDAOKSKORLEKELEKAAEVPTEYTL
VPSIDTTHKNGKREFLEOVNTKAFGERETLEKANNPTIOHGEVATIGNGSGKTTLL
NIIIGDTEARSGSVWSPSANITGLTCEYFPLPLEQPEELFENETFKARGHYONIMRH
LGPAAQVTEPRPKHMSGCEYKIKLAVYLLEEDVLLIDEPYHLDLPSREDETLIS
QYSLTAVSHDRVLEKTYNSKIVISNGLERGLNDVSENRERRELLKLETEROE
VLKLSFMTPNDRGTKELDQAFNELTKIKELHQRKD"
complement(3428. .5450)
/gene="ydgq"
complement(5428. .7089)
/gene="ydgq"
complement(5678. .5703)
/gene="ydgq"
complement(5713. .7089)
/gene="ydgq"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=1
/protein_id="CAB12381.1"
/db_xref="GI:2632875"
/db_xref="SWISS-PROT:P96704"
/translation="MTDMDKNTKNTQOQLQRLGNRIQYLAIGALIGTGLFSGSKS
IHRAGSTILFAVMTGITICFLIMKSLGSLNSNHSYDFVQDVLGMAAFITMT
YMCWSTIMADVLAVGLTYQWLPVGPWPVGLALITLLINKLAVKILFGLSEWVE
ALIKVATIALIALIVGLVMIKFGFSTSSGVFTNLMWGGGLPFGNHGFLISFQVFE
APVQELVGLTIDENPEKVIKPAIKNIIPVRIEFLYGLALVIMSYIPMDIINPSES
PFOVQVFAVGIYGAASIIINPVLTLSAASACSAVFTSRVYSLANDHNPESMAKIT

alignment_scores: Quality: 958.50 Length: 313
 Ratio: 3.804 Gaps: 1
Percent Similarity: 80.511 Percent Identity: 55.591

alignment_block:
US-09-462-846-2 x BSUB0004 ..
Align seg 1/1 to: BSUB0004 from: 1 to: 213190

3 rhrcluproleuphepelysprovalpelysgluarqtlletrpelyl 19
30827 ACGCATCCATTATTTTAAAGCCCTGCTTTAAAGAAAGACTATGCGAGG 30876
19 yhrhaleualaas...pheglytyrthrleprosergluarqthrg 35
30877 GACGAAGCTCGTGACGCTTTGGCTACGCAATACCTGCAAAAAACAG 30926
35 lyglucysrpalaphealajshisglinsnglyginservalvalgin 51
30927 GTGATGCTGGGCGCTTCTGCACATGCGCATGCTGCTGTGTAANA 30976
52 Aanglymettyrlysglypetherleusergluhleutrpeluhishisar 68
30977 AATGACCGCTGGCGAGGAAGACACTGATCAAGTAAAGAAATATATCC 31026
68 ghlsaupheglygluenglugliyaapargpneproleuleuthrlysi 85
31027 ACAGATATTCGGGTTCCGATGCTAAGGTTGCTGCTGCTGCTGTAAC 31076
85 leleuaspalaaapclnaspheuservalglnvalhisprohnaapglu 101

US-09-462-846-2 x AE006310 ..

Align seg 1/1 to: AE006310 from: 1 to: 10607

```

4  GIUProleuPhepelysProValPheLysGluArg1LeuPrgLys1 20
   |||..... |||..... |||..... |||..... |||.....
7717 GAACCATGTTGTTGACCTGTTGCAAGAAATAATCGGGCGCGCA 7766
20  rAlaLeuAlaAspPheGlyTyrThrIleProSerGlnArgThiGlyL 37
   |||..... |||..... |||..... |||..... |||.....
7767 CCATTTAAAGAGTTGGCTATGATTTGCCATGACAAAGTTGGTGAAT 7816
37  ysrPAlaPheAlaAlaHisGlnAsnGlyGlnSerValAlaGlnAsn 53
   |||..... |||..... |||..... |||..... |||.....
7817 ATGGGCTATTTGCTGCTATCCATCCATGCTGTGCAACAATTCCTAA 7866
54  MetTyrLysGlyPheThrLeuSerGlnLeuTrpGlnHisArgHis 70
   |||..... |||..... |||..... |||..... |||.....
7867 GAATTTAAAGTCAAAAGCTTGACCATTTATGCAAGTCAAGCCGCAAT 7916
70  uPheGlyGlnLeuGlyLysPArgPheProLeuLeuThrLysIleLeu 87
   |||..... |||..... |||..... |||..... |||.....
7917 GTTGGGATGATGATAAAAGATTTTCCCTTACATACTAAATTTTATG 7966
87  sPAlaAspGlnAspLeuSerValGlnValHisProAsnAspGluTyr 103
   |||..... |||..... |||..... |||..... |||.....
7967 ATGCCAATGACTGGCTTTCTGTCGCAATGCTCATCCAGATGAATATG 8016
104 AsnIleHisGlnAsnGlyLysGlyLysThrGlyCysTrpTyrIle 120
   |||..... |||..... |||..... |||..... |||.....
8017 CAAAAACATGAA...GGTGAACCTTGCAAAAACATGATGTTGATCAT 8063
120 eAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsnAlaThr 137
   |||..... |||..... |||..... |||..... |||.....
8064 TTCAGCTCAACAGCCGCTGATATCTATGACATAAAGCAATTAATC 8113
137 ysgLglnLeuThrThrMetIleGluArgGlyLysTrpAspGluLeu 153
   |||..... |||..... |||..... |||..... |||.....
8114 GTCAAGATTTAGCAGAAATGATTAATCTGATGATGATGATGATGAT 8163
154 ArgAlaGlyValLysValLysProGlyAspPhePheValProSerG 170
   |||..... |||..... |||..... |||..... |||.....
8164 CGTAAAGGTAAGTGAACAGGGATTTCTTCATGTTCCGTCAGAGAAC 8213
170 rValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGlnGln 187
   |||..... |||..... |||..... |||..... |||.....
8214 AATGACGCAATCGGTGCTGGAATGTTATTTCTTGAACACAAACAT 8263
187 eAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspAlaGlu 203
   |||..... |||..... |||..... |||..... |||.....
8264 CTGATACAACTTACCGTGTATATGATTTCCACCTTAATAAGTCAAG 8313
204 LysLeuArgGluLeuHisLeuLysSerIleGluValIleGluVal 220
   |||..... |||..... |||..... |||..... |||.....
8314 AATCTAGCTGATTTACATATTCACAAACATTTATGATTTGAATATTC 8363
220 o.....serIleProGluArgHisThrValHisGluGlnIleGlu 235
   |||..... |||..... |||..... |||..... |||.....
8364 GGGCACAAGATCCCTGAA...AATCAATTTAAACGAAAAAATTCCTG 8410
235 sPLeuLeuThrThrThrLeuIleGlyCysAlaTyrPheSerValGly 251
   |||..... |||..... |||..... |||..... |||.....
8411 ATGCAAAATTTACAACTTTGTGAATCAAAATTTCTTGATGTTTATAA 8460
252 TrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLe 268
   |||..... |||..... |||..... |||..... |||.....
8461 TGGCAAAATTCATGCTGACCAATGAAATTTACAAATTCCTGATTAAC 8510
268 uIleSerValIleGlyGlyGlyArgMetIleSerGlyCysLysVal 285
   |||..... |||..... |||..... |||..... |||.....
8511 AGTTCTGATCTGATGCTCAAGGAAATTAACAATGATGATGAATGAAT 8560
285 yPProPheLysGlyAspHisMetLeuProTyrGlyLeuGlyGlu 301

```

||||| |||..... |||..... |||..... |||.....

8561 ATCCAGTTGAAAAAGAGCTCATTTCTATCTTACCAAGCAATGAAAAA 8610
 302 PheLysLeuGlnGlyTyrAlaGlyCysIleValSerHis 314
 ||| ||| ||| |||
 8611 TGAATTTGCTGCTCAATTAAGAAATTAATGCGACCAAT 8649

seq_name: gb_htg25:SPNEU1904

seq_documentation_block:

LOCUS SPNEU1904 147006 bp DNA HTG 16-DEC-2000
 DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
 in ordered pieces.

ACCESSION AL449926.1 GI:11545151
 VERSION AL449926.1 GI:11545151

KEYWORDS HTG, HTGS, PHASE2.
 SOURCE Streptococcus pneumoniae.

ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 Streptococcus

REFERENCE 1 (bases 1 to 147006)
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Calera,F., Polissi,A.,
 Humbert,Y., Friedl,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
 Francesco,M., Buehl,G., Feger,G., Garcia,E., Peltsch,M. and
 Garcia-Bustos,J.F.

JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
 Severo Ochoa 2, 28760 Tres Cantos, SPAIN

COMMENT * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 Location/Qualifiers

Source 1..147006
 /organism="Streptococcus pneumoniae"
 /serotype="19f"
 /db_xref="taxon:1313"
 /clone="G54"

BASE COUNT 44762 a 26837 c 32084 g 43281 t 42 others

ORIGIN

alignment_scores:
 Quality: 839.00 Length: 316
 Ratio: 3.467 Gaps: 5
 Percent Similarity: 76.582 Percent Identity: 53.797

alignment_block:
 US-09-462-846-2 x SPNEU1904 ..

Align seg 1/1 to: SPNEU1904 from: 1 to: 147006

3 ThrGluProLeuPhepelysProValPheLysGluArg1LeuPrgLys 19
 |||..... |||..... |||..... |||..... |||.....
 79069 TCAGAACCATTAATTTTACATCTGATGACAAATAAATACTGGGGTGG 79118

19 yThrAlaLeuAlaAsp...PheGlyTyrThrIleProSerGlnArgThi 35
 |||..... |||..... |||..... |||..... |||.....
 79119 AGCCAAAGCTAGCTAGTATGCTTGGCTACGACATCCCAAGTCAAAAAATTC 79168

35 yGlnCysTyrPAlaPheAlaAlaHisGlnAsnGlyGlnSerValAlaGln 51
 |||..... |||..... |||..... |||..... |||.....
 79169 GAGATATTTGGGCGCATCTCAGCCCATCCAAATGAGTCTTAAGTTGCC 79218

52 AsnGlyMetTyrLysGlyPheThrLeuSerGlnLeuTrpGlnHisArg 68
 |||..... |||..... |||..... |||..... |||.....
 79219 AATGGCTTTCAGGAGGAAAGATCTTGCTTGTGATCGGAAACCG 79268

68 gHisLeuPheGlyGlnLeuGlyLysPArgPheProLeuLeuThrLys 85
 |||..... |||..... |||..... |||..... |||.....
 79269 TGAATTAATTTGGCAATGCTCCAGAACTGATATTCACACTTTGACCAAGA 79318

85 leuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 101

```

|||||..... |||||||||||||||||||||
79319 TCCCGAGTCCCAAGACTGCTCAGTGTCCAAATTCACCACGACGATGCT 79368
102 TTAAlaasnlleHlsGluasnGlyGluLeuGlyLysThrGluCysTrpTy 118
|||||..... |||||||||||||||||||||
79369 TATCGACTCGAGCATGAA...GGCGAACTCGCAAAAGAAAGATGCTGTA 79415
118 TlellaAspCysGlnLysAspAlaGluIleIleTyrglyHlsAsnAlaT 135
|||||..... |||||||||||||||||||||
79416 CATATCGCAGCGATGACATGATGATGATGATGATGATGATGATGATGAT 79465
135 hrTrlysgLugluLeuThrThrmellleGluArgGlyGluTrpAspGlu 151
|||||..... |||||||||||||||||||||
79466 AGTCAAAAGAAAGAACTCCGCCAAATTCAGGACAAAGAACTGGATGAC 79515
152 LeuLeuArgArgValLysValLysProGlyAspPhePheTyValProse 168
|||||..... |||||||||||||||||||||
79516 TTGTGGACCAAGTCCCTGTTAAGGCTGAGATTTCTGTATGTACCAAG 79565
168 TGLYThrValHlsAlaTleGlyLysGlyLysAlaLeuAlaLeuGluThrGlu 185
|||||..... |||||||||||||||||||||
79566 CGGACTATGACACGCTATCGGGCGGATATTAATCTTGAAACCACGAC 79615
185 lnaAsnSerAspThrThyArgLeuTyraSPTyrAspArgLysAspAla 201
|||||..... |||||||||||||||||||||
79616 AGTTCAGCATACCACTATCGGCTATGATGATGATGATGATGATGATGAT 79665
202 GluGlyLysLeuArgGluLeuHlsLeuLysSerLleGluValLleG 218
|||||..... |||||||||||||||||||||
79666 AAAGCAACTGCGCTGAACTTCACTTGAAATAATCCATGATGCTGTA 79715
218 uVal.....ProserLleProGluArgHlsThrValHlsHlsGluGln 233
|||||..... |||||||||||||||||||||
79716 CATGTGTGACCCACCAATAGCCGCTCTTAACCTTT.....AAAG 79756
233 legLnaAspLeuLeuThrThrThrLeuLleGlyLysAlaTyPheSerVal 249
|||||..... |||||||||||||||||||||
79757 CAGATGATTTGCGCTCCACCTCTCTGTATCATGATTTCTTCGCGAGTT 79806
250 GlysTrpAsnLeuSerGlySerAlaSerLeuLysGlnLysProPh 266
|||||..... |||||||||||||||||||||
79807 TACAAGGGGAAATTAAGTGAAGAAAGTGAAGAAAGTGAAGTGAAGTGA 79856
266 eleuLeuLleSerValLleGlyLysGlyLysArgMetLleSerGlyLys 283
|||||..... |||||||||||||||||||||
79857 CAGCTTATGATGCTCTTACGTGCTCAAGTCAAGTCAAGTCAAGTCAAGT 79906
283 yrValTyTrpPheLysGlyAspHlsMetLeuLeuProTyrglyLeu 299
|||||..... |||||||||||||||||||||
79907 AAAATTTTCAATTCAAAAGGAGACACTTATCCACCAAGTGAATGTT 79956
300 GlysLysPheLysLeuGlyLysAla...GlyCysLleValSerHls 314
|||||..... |||||||||||||||||||||
79957 GAACCTTGACTGCTGGAAGGCAAGGTTTGAATGATGTTAGCAT 80004
seq_name: gb_hlg12:AC027136
seq documentation block:
LOCUS AC027136 49897 bp DNA HTG 30-JAN-2001
DEFINITION Staphylococcus aureus clone sabac-106, WORKING DRAFT SEQUENCE, 1
ORDERED PLACES.
AC027136 GI:8567822
VERSION AC027136.5
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
1 (bases 1 to 49897)
REFERENCE 1. (bases 1 to 49897)
AUTHORS Loh, P., Qi, S., Ray, L., Ford, B., Iandolo, J., and Roe, B. A.
TITLE Staphylococcus aureus BAC Clone sabac-106
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 49897)

```

```

AUTHORS Loh, P., Qi, S., Ray, L., Ford, B., Iandolo, J., and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jan 15, 2000 this sequence version replaced gi:8225121.
NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 49897: contig of 49897 bp in length.
FEATURES
location/Qualifiers
source
1. 49897
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
/clone_lib="ou staph library"
BASE COUNT 15753 a 9237 c 7267 g 17640 t
ORIGIN
alignment_scores:
Quality: 819.50 Length: 312
Ratio: 3.415 Gaps: 1
Percent Similarity: 76.923 Percent Identity: 50.321
alignment block:
us-09-462-846-2 x AC027136
Align seg 1/1 to: AC027136 from: 1 to: 49897
3 ThrGluProLeuPhePheLysProValPheLysGluArgLleTrpGly 19
|||||..... |||||||||||||||||||||
35576 ACTATGCGATTAATTTTACAAACCCGTTTAAAGACGAATTTGGGCGG 35625
19 yThAlaLeuAlaAspPheGlyTyThrLleProSerGlnArgThrGly 36
|||||..... |||||||||||||||||||||
35626 TCATGCAATTAAGCAATTTACCTGATATGCCATGAACAACAGCGTG 35675
36 luySTrPAlaPheAlaAlaHlsGlnAsnGlyGlnSerValLleGlnAsn 52
|||||..... |||||||||||||||||||||
35676 AATGCTGGCAATTTCTGCACATCCAAATGCTCTATATATATATCAAT 35725
53 GlyMetTyTrpPheLysGlyPheThrLeuSerGluLeuTrpGluHlsArgH 69
|||||..... |||||||||||||||||||||
35726 GGTCCATGACAGGATATGACCCATGATCAACTTGGTCACACATCGTGA 35775
69 sleuPheGlyLysLeuGlyLysAspArgPheProLeuLeuThrLysLle 86
|||||..... |||||||||||||||||||||
35776 ATTAATCGATATGATTCACGAGATGTTCCCTGCTAATCAAGTAT 35825
86 euAspAlaAspLysPheLysSerValGlnValHlsProAsnAspGluTy 102
|||||..... |||||||||||||||||||||
35826 TAGATGCCAATGATTAATATCTTCAAGTCAACCCAGATGATGATC 35875
103 AlaAsnLleHlsGluasnGlyGluLeuGlyLysThrGluCysTrpTy 119
|||||..... |||||||||||||||||||||
35876 GCTTAAACAGAA...GGCGAACTGAGTAAACACAGAAAGTTGGGTAT 35922
119 eLlAspCysGlnLysAspAlaGluIleIleTyrglyHlsAsnAlaThr 136
|||||..... |||||||||||||||||||||
35923 TTATGATGACACCCAGGCTGTAATATATATATATATATATATATATAT 35972
136 hrTrysGluGluLeuThrThrmellleGluArgGlyGluTrpAspGlu 152
|||||..... |||||||||||||||||||||
35973 ATAAACAGACTTAATAGACATGATTAATATATATATATATATATATAT 36022
153 LeuArgArgValLysValLysProGlyAspPhePheTyValProSerG 169

```



```

36023 TTCAAACGTAACCTGTAAGCCCGGTGACTTTATTATGTTCCCTGCG 36072
169 yThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGlnGln 186
36073 CACTGCCATGCCATGTGTCGGGTATTTTATTATTTAGACACAAACAT 36122
186 snSerAspThrThrTyArgLeuTyArgPtyArgPtyArgLysAspAlaGlu 202
36123 CCYCCGATCGACATCCGATTTATGACTATGATGACGAGATCAAAAT 36172
203 GlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleGluVal 219
36173 GGTCAATTCGGTGAATTTGCACCTTAGCAAGAAAGATGATGATGAAT 36222
219 lProSerIleProGluArgHisThrValHisHisGluGlnIleGluAspL 236
36223 AGGCAATCATGAACCAAAATGACTCTCTATACACACTCAGATGACACAC 36272
236 euLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLysTrp 252
36273 ACACGTGTAACATAATTGTAGATATCAATATTTTGCAGATATATAATG 36322
253 AsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeu 269
36323 GATATTCATGGAGATTAAATTCAAATCAATAACACATGCTTATGCTTAG 36372
269 eSerValIleGluGlyGluGlyArgMetIleSerGlyGluTyrValTyrP 286
36373 TACAAATTATGATGATGACTGTACTTTAACAGTAACGATAGCACTATG 36422
286 ropheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGluPhe 302
36423 ACATCCAAAGAGGTCTAGCTTATTTTAACTACGGAGAGAGAGATATT 36472
303 LysLeuGluGlyTyrAlaGluCysIleValSerHis 314
36473 CACTTGTGCGTAATATATAGTCTATAGTAAGTCAAT 36508
```


Fri May 18 10:38:05 2001

us-09-462-846-2.p2n.rst

Page 3

BP 191 91006 Evry cedex - France
Email: segret@genoscope.cns.fr, Web : www.genoscope.cns.fr
Location/Qualifiers
1. 830

```

/organism="Homo sapiens"
/db_xref="takon:3606"
/clone="CS0DF019YF01"
/clone_id="PL1_PL013_Brn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="Bnuhp"
/lab_host="Bnuhp"

```

/note-**Origin:** Fetal brain; **Vector:** pCMWSP6; **1st strand cDNA** was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMWSP6 vector. Library was constructed by Life Technologies. **Contact:** Feng Liang Liang Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville Maryland 20850, USA Fax: (1) 301 610 8371 Email: liangliang@litech.com URL: <http://fulllength.invitrogen.com>

BASE COUNT	227 a	188 c	210 g	205 t
ORIGIN				

```
alignment_scores:
  Quality: 93.50
  Ratio: 0.760
  Gaps: 11
Percent Similarity: 49.398
Percent Identity: 18.876
```

```
alignment_block:
US-09-462-846-2 x AL537077 .
```

Align seg 1/1 to: AL537077 from: 1 to: 830

[illegible]

```

445 CTGCATCGCTGTTTCCCTTGGAAACAACATCTCTGATCACAATT...CGCAT 47
147 YGLTTPASPGILuLeuLeuAArgValLysValLysProGlyAspPhe 164
472 AGTTTTCCAAAGCGATTTTCCGCTTTGGGTGGTATCAACAAGCTGCA 521
164 heyr.....ValProser.....GlyThrVal 171
522 TATTTATCCAAATGTGTGCATTAATACACAGTCGCTTTATGGAAAGCTG 571
172 HisAlaIleuLysLysLysLeuAlaLeuGluThrGlnAsnSerAs 189
572 GAACCTACACCCAAATCCTTATTCAGCCAAAGACAGCCGACCAAGA 621
188 pThhThTy.....ArgLeuTyAspTyra 197
622 GAATTCATTTTCAAAAGCTGATGCTGAATATAAAAAACTTCATAGTATG 671
197 sPaGlyAspAlaGluGluLysLeuAArgLeuGluLeuHisLeuLysLys 212
672 GAAACA...GACCAGAAAAGGATGATGAAAGAAAGACTTCAACCAAGCA 715

```

seq_name	qb_est8	seq_documentation	block	675 bp	EST	06-DEC-X000
LOCUS	BF492790					
DEFINITION	AF01206 5prime AT Drosophila melanogaster adult testes polyA7					
	Drosophila melanogaster cDNA clone AT01206 5 similar to CG1161					
	EFan0011661 last updated:000321, mRNA sequence.					

VERSION	BF492790.1	GI:11576091
KEYWORDS	EST.	
SOURCE	fruit fly.	

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 675)

REFERENCE

REFERENCE AUTHORS

B. Carlson, E. Crampe, M. Chavez, C. Chew, M. Huet, V. Fiala,
D. Frisze, E. George, R. Gonzalez, M. Guerin, H. Harris, N. Li, P.
Liao, G. Miranda, A. Mista, S. Mungall, C. J. Nuno, J. Paley, J.
Park, S. Parada, V. Pounanevong, S. Wan, K. Yu, C. Zelniker, S.
Lewis, S. E. and Rubin, G. M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Other-ESTs: AY01206..3prime

contact: stapletron, m.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd., Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: hit@www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic sequence AE003525, nit genomic sequence AE003472
Plate: AT_12 row: A column: 6
high quality sequence stop: 595.

FEATURES

SOURCE

```

/organism="Drosophila melanogaster"
/bb_xref="taxon:7227"
/clone="AT01206"
/clone_lib="At Drosophila melanogaster adult testes poNB7
/sex="male"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="DH5-alpha or DH5-alpha Tona as per database (AT
121 on are in Tona cells)"
/notes="Organ: ADULT testes; Vector: poNB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into poNB7. Plasmid cDNA library."

```

ORIGIN

Alignment_scores: Quality: 93.00 Length: 200
 Ratio: 0.959 Gaps: 8
 Percent Similarity: 48.500 Percent Identity: 24.000

Alignment block:
 US-09-462-846-2 x BF492790 ..

Align seg 1/1 to: BF492790 from: 1 to: 675

```

146 ArgGlyGluTrpAspGluLeuLeuArgValysValysProGlyAs 162
|||||
51 CGTGCATCAACGACGACACCTGAGAGCATTT.....GGCAA 88
162 PhePheTyrValProSerGlyThrValHisAlaIle.....GlyLysG 177
|||||
89 CATGTTCTCGACTCTCTCCGCCGCCGAGCAAGTTGAGACACACAAAG 138
177 LysLeuAlaLeuGluThrGlnGlnAsnSerAspThrThrTyrArgLeu 193
|||||
139 GCATCCTTCGCAATCACTGCGCCACGACGACGACAGTGTGAGGACAAAGTG 188
194 TyrAspTyrAspArgLysAspAla.....G1 202
|||||
189 GCGGACTGTGGCTGGGCGGAGGCTTCGCTTCGGTTCGCTCTTAAGGA 238
202 uGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleGluV 219
|||||
239 GGGG.....ATCCACGTTCCGACTATCCGACGAGTGATGGAGC 276
219 aLProSerIleProGluArgHisThrValHisHisGluGlnIleGluAsp 235
|||||
277 GCGGACACTTTTGGCACCAGGACACCACTCTCTGACACAGTCGAGGAGAC 326
236 LeuLeuThrThrThrLeuIleGluCysAlaTyr.....Ph 247
|||||
337 AAGGTGGTGTACAACTCACTGACACCTCTATCCGACACAGGCCCTTA 376
247 eSerValGlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGln 264
|||||
377 CTCGCTCTCCAAACAGTCCCTCTGAGAGTGGCGGCTCGGCTTCGAGC 426
264 ySPProPheLeuLeuIleSer.....ValIleGluGlyGluGlyArg 277
|||||
427 ACGGCTACTCCATGGCCAGTCCCATGCTCTGTGATGTGGGAGAGCCAG 476
278 .....Met11 279
477 TTGGGACACTTTCGAACACGCGGACGATCATCATGACAGCTTCATGAC 526
279 eSerGlyGluTyrValTyrProPheLysGlyAspHisMetLeuLeuP 296
|||||
527 CAGGCGGGAACCAAGTGGGTGGCGGACGCGGAGTCTCATGTTCGTCGC 576
296 rGlyr.....GlyLeuGlyGluPheLysLeuGluGlyTyrAlaGlu 309
|||||
577 CACACACACATGAGAGGCGATGGGCCCGGACGACACTTTCGCGCGCATTCAG 626

```

seq_name: gb_estf5:BE567646

seq_documentation_block: 869 bp mRNA EST 15-AUG-2000
 LOCUS BE567646
 DEFINITION 6013402933F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3682654 5',
 mRNA sequence.
 ACCESSION BE567646
 VERSION BE567646.1 GI:9811366
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

REFERENCE
 1 (bases 1 to 869)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.C.Strausberg@nih.gov

Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
 Plate: LNCM366 row: 1 column: 23
 High quality sequence stop: 481.

FEATURES

Location/Qualifiers
 1..869
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3682654"
 /clone_id="NIH_MGC_53"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Bladder; Vector: pDNM-LIB (Clontech); Site_1: Still (ggcgcctcggcc); Site_2: Still (ggcctatagcc 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGACC-3' and 3' adaptor sequence: 5'-ATCTGAGAGCGGCGGCGGAGATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT

196 a 228 c 215 g 230 t

Alignment_scores: Quality: 93.00 Length: 241
 Ratio: 0.809 Gaps: 11
 Percent Similarity: 47.718 Percent Identity: 20.332

Alignment block:
 US-09-462-846-2 x BE567646/rev ..

Align seg 1/1 to reverse of: BE567646 from: 1 to: 869

```

58 PheThrLeuSerGluLeuTrpGluHisHisArgHisLeuPheGlyGln 74
|||||
833 TATGAAATAGCTGTGATGTGGCACCCACCC.....TCATGAGA 796
74 uGluGluAspArgPheProLeuLeuThrLysIleLeuAspAlaSerAla 91
|||||
795 AGAGAGGGGGTTTTCCTCCACAGACACAGATGTGTGGGCGCGCCGCC 746
91 sPLeuSerValGlnValHisProAsnAspGluTyrAlaAsnIleGlu 107
|||||
745 GGGGCCCCCTTCAGTGTGCCA.....CACGAA 717
108 AsnGlyGluLeuGlyLysThrGluCysTrpTyrIleIleLeuAspCysGln 124
|||||
716 CAACACACTGTGTGGCGTTTTCATTT.....TCATATTA 682
124 sAspAlaGluIleIleTyrGlyHisAsnAlaThrThrLysGluGluLeu 141
|||||
681 A...AAGAAATGCCCATTAAGCGTGGGCGCCCAATATTAACGAGTTA 635
141 hTrhMetIleGlu.....ArgGlyGluTrpAspGluLeuLeuArgArg 155
|||||
634 CCGTTTATTGACCCCAACAAAGAAATGGGGAAG..... 597

```

```

156 VallyVallystroglyAspPhePheValProsergylThrValH 172
159 .....
596 .....GGGACATTAAGAAATTAAGAGGACCACTACTA 562
172 salalleglysglyleuvalaleuGlutThrglnGlnAsnSerSPT 189
561 CCGCTTAGGTAGTACAGACTGTGGTGGGAACCCA.....ACTGACA 518
189 hTThrTyrArgLeuTyrAspTyrArgValAspAlaGluGlyLysLeu 205
517 CGCTGTCTTAACCT.....GAGAAAGTAATCTTGGTCTTTT 480
206 ArgGlu.....LeuHisLeuLysSerLeuGluValIleGluVal 219
479 AAGAAACAAAGATCTTGGAGGTAAAGAACAGAGGTCCCTTGTCAAGT 430
219 lProserllepTroglyArgHisThrValHisGlnGlnIleGluAsp 236
429 AGACACCCCTCCCTT.....
236 euleuThrThrLeuIleGlyAlaTyrPheSerValGlyLysTyr 252
413 .....GTCGTGTGGACACACAACTGG 390
253 AsnLeuSerGlySerAlaSerLeuGlnGlnLysProPheLeuLeu 269
389 AATCTGCATCGCGCAGCTCCAAACATG..... 360
269 eSerValIleGlyGlyGlyArgMetIleSerGlyLysTyrValTyr 286
359 .....GGGCAAGGTGGTCTGTTCATATGATGTGTGCCCC 323
286 ropheLysGlyAspHisMet 293
322 CGTCACAAATCGGATCTGATCTG 300
seq_name: gb_gss21.AZ432549

```

seq_documentation_block:

LOCUS AZ432549 710 bp DNA GSS 03-OCT-2000
DEFINITION 1M0218601F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0218601 F, DNA sequence.

ACCESSION AZ432549
VERSION AZ432549.1 GI:10556562

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddun@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0218 row: 6 column: 01
Seq primer: CGTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 710.
location/Qualifiers
1..710
/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0218601"
/clone_id="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain X110-gold, T1-resistant, F-"  

/note="Vector: PMD42uv; Purified genomic DNA from Mus  

musculus C57BL/6J (male) was obtained from the Jackson  

Laboratory Mouse DNA Resource  

(http://www.jax.org/resources/documents/dnares/). The DNA  

was hydrodynamically sheared by repeated passage through a  

0.005 inch orifice at constant velocity. The sheared DNA  

was blunt end-repaired with T4 DNA polymerase and T4  

polynucleotide kinase. Adaptor oligonucleotides were  

ligated to the blunt ends in high molar excess. The  

adapted DNA was purified and size-selected for a 9.5 to  

10.5 kb range using preparative agarose gel  

electrophoresis. Vector DNA was prepared from a derivative  

of PMD42 (g1147321149b/AF129072.1), a copy-number  

inducible derivative of plasmid R1. The vector was ligated  

with adaptors complementary to the insert adaptors and  

purified. The sheared, adapted mouse DNA was annealed to  

adapted vector DNA, and transformed into  

chemically-competent E. coli X110-gold (Stratagene) cells  

and selected for ampicillin resistance."
BASE COUNT 249 a 129 c 182 g 150 t
ORIGIN

```

alignment_scores:
Quality: 92.50 Length: 102
Ratio: 1.779 Gaps: 2
Percent Similarity: 50.980 Percent Identity: 25.490

alignment_block:
US-09-462-846-2 x AZ432549 ..

Align seg 1/1 to: AZ432549 from: 1 to: 710

```

38 TTPAlaPheAlaIleHisGlnAsnGlyLysSerValValGlnAsnGlyme 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 TCGAGAGTGGTTCAGGATAGAACAGGCCCTAAGCACTTCAAAATGATCC 387
54 tTyrLysGlyPheThrLeuSerGlnLeuTyrGlnHisHisArgHisLeu 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 AATTAAGGAGTGTAAATTCACAGCTTGAAGCTGAGAAAAAGTGTGT 437
71 heGlyGlnLeuGlyAspArgPheProLeuLeuThrLysIleLeuAsp 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 GTGAAAGGCGAGAACAGCACTTCCAGAGACAGCAAAAGTTTTCAG 487
88 AlaAspGlnAspLeu..... 92
488 TTTCGAAGTATGTCTAAATCCCTTAAAGAGCAACAAAGCACTTGG 537
93 .....SerValGlnValHisProAsnAspG 101
538 GTACAGGTAGATGACACAGAGAGCCGATGCGAGTCC.....CCAC 578
101 lutyAlaAsnIleHisGlnAsnGlyLysLeuGlyThrGlyCysTyr 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
579 AGGCAACCAACGTTCCAGAGATGGGCAAGTGGCACAGCACACTGCTGG 628
118 TyrIle 119
|||:
629 TTGCTG 634
seq_name: gb_est88:BF489340

```

seq_documentation_block:

LOCUS BF489340 577 bp mRNA EST 06-DEC-2000
DEFINITION AT25326.Sprtime AT Drosophila melanogaster adult testes P0787
Drosophila melanogaster cDNA clone AT25326 5 similar to CG7934:

ACCESSION	FBan0007934	last_updated:000321,	mrna sequence
	BF489340		

VERSION BF489340.1 GI:11572641
KEYWORDS ECM

NEIMOLDS
SOURCE
ORGANISM

REFERENCE AUTHORS

BFan0007934.last.updated:000321, mRNA sequence.
 BF489340
 BF489340.1 GI:11572641
 EST.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 577)

TITLE	Berkeley Drosophila Gene Collection Project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Stapleton, M.

Brady
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA.
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit@genomic.sequence.AE003525 hit genomic sequence AE003472
Plate: AR.253 row: C column: 2
High quality sequence stop: 575.

FEATURES

```
/organism="Drosophila melanogaster"  
/db_xref="taxon:7227"
```

/clone="AT55326"
 /clone_1fb="At Drosophila melanogaster adult testes pOTB7"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_host="DH5-alpha or DH5-alpha Tona as per database (ATP121 on are in Tona cells)"
 /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."

BASE COUNT	1
ORIGIN	

114 a	187 c	171 g	105 t
-------	-------	-------	-------

alignment_scores:

Quality:	92.00	Length:	185
Ratio:	1.011	Gaps:	7
Percent Similarity:	49.189	Percent Identity:	24.324

```
alignment_block:
```

Align seg 1/1 to: BF489340 from: 1 to: 577

[illegible]

```

217 egLVaIProSerLleProGLuATGhIsthrValhIshsgLugInIleG 234
    ||||| :||||| :||||| ||| |||||:|||||
220 GGAGCGGGCGACCTTTTGGCAGCGGCACACAGCTCCGACACCGACAGTGG 265
    ||||| :||||| :||||| :||||| :||||| :|||||
234 LuAspLeuLeuthrThrThrLeuIleLcylsAlaIyr..... 246
    ||||| :||||| :||||| :||||| :||||| :|||||
270 AGGCAAGAGGGGGTACACACTACGTGGACCACTGTATCCGACCGAGCG 319
    ||||| :||||| :||||| :||||| :||||| :|||||
247 ...PheSerValGlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGI 262
    :||||| :||||| :||||| :||||| :||||| :|||||
320 CCGTACTCGGTCTCCACAGAGCTCCCTGTCCGAGTGTGCCGTCTTGGCGTT 365
    :||||| :||||| :||||| :||||| :||||| :|||||
262 ngLInLysProPheLeuLeuIleSer.....ValIleGluGluGluG 276
    :||| :||| :||| :||| :||| :|||
370 CGAGACAGCGTACTCTCATGGCCAGTCCCAATGCTGTGGTCAATGTGGAG 419
    :||| :||| :||| :||| :||| :|||
276 LysArg..... 277
    ||||| :||||| :||||| :||||| :||||| :|||||
420 GACAGCTTTGGGGACTTCTGCACACAGCGCGAGTGCATATCGACACGTTG 468
    :||||| :||||| :||||| :||||| :||||| :|||||
278 MetLleSerGlyGluTyrValTyrProPheLysLysGlyAspHisMetLe 294
    :||| :||| :||| :||| :||| :|||
470 ATACACACAGCGCGGAAACCAAGTGGGTGGCGGACATCCGGAGTGTGATGTT 519
    :||||| :||||| :||||| :||||| :||||| :|||||
294 uLeuProTyr.....GlyLeuGlyLupPheLysLeuGluGluTyrA 308
    ||||| :||||| :||||| :||||| :||||| :|||||
520 GCTGGCCACACAGATGAGGGCGATGGGCCCTTGAGCAGCTTTCGGCGCGCA 565
    :||||| :||||| :||||| :||||| :||||| :|||||
308 LysIlu 309
    |||
570 TTTGAG 574

seq_name: gb_est71:BE204501

```

seq_documentation_block:	551 bp	EST	05-SEP-2000
LOCUS	BE204501	mRNA	
DEFINITION	Sequence 3917177	KV0	Medicago truncatula cDNA clone pkv0-16010, MRN
ACCESSION	BE204501	sequence	
VERSION	BE204501.1		
KEYWORDS	EST.		
SOURCE	Barrel medic.		
ORGANISM	Medicago truncatula		

REFERENCE AUTHORS

Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.

TITLE	ESTs from uninoculated seedling roots of <i>Medicago truncatula</i>
JOURNAL	Unpublished (1999)
COMMENT	Contact: VandenBosch K

College Station, TX 77843-3258, USA

FEATURES
Source

```

/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="PKV0-16010"
/clone_lib="KV0"
/tissue_type="Seedling roots"
/dev_stage="Immediately prior to inoculation with"

```


Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Celniker, S.,
Lewis, S.E. and Rubin, G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
COMMENT
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AE003525; hit genomic sequence AE003472
Plate: AT.264 row: G column: 3
High quality sequence stop: 681.
Location/Qualifiers

FEATURES

source

1. 708
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT28475"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="DH5-alpha or DH5-alpha Tona as per database (AT
121 on are in Tona cells)"
/note="Organ: Adult testes; Vector: pOTB7; Site_1: EcORI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 148 a 227 c 203 g 130 t
ORIGIN

alignment_scores:
Quality: 91.50 Length: 193
Ratio: 0.984 Gaps: 8
Percent Similarity: 48.187 Percent Identity: 23.834

alignment_block:
US-09-462-846-2 x BF491677 ..

Align seg 1/1 to: BF491677 from: 1 to: 708

155 ArgValLysValLysPro..... 160
125 CGCTAAACTGTGCCCGCATGACGACGACACCTGACAGCAT 174
161 .GlyAspPhePheTyrValProSerGlyThrValHisAlaIle.....G 175
175 TGGCAGCATGTTCTCGACTCGCGCGGACGACAGAGTTTCAGACAC 224
175 LysGlyIleLeuAlaLeuGlnThrGlnGlnInsSerAspThrTyr 191
225 ACAAGGCGATCTTCGCATCTGCGCCAGCGACGACAGATGTCAGAC 274
192 ArgLeuTyrAspTyrAspArgLysAspAla..... 201
275 AAAGTGGCGACTGTGCTGGCGGAGGCTTCGCTCGCTCGCTCT 324
202GluGlyLysLeuArgGlnLeuHisLeuLysSerIleGluValI 217
325 TAAGGAGAGGC.....ATCCAGGTTGACATATCCGGCAGGATG 362
217 IeGluValProSerIleProGluArgHisThrValHisGlnGluGln 233
363 TGGAGCGCGGACCTTTCGACCGGACGACGATCTCGACACGACGTCG 412
234 GluAspLeuThrThrThrLeuIleGluCysAlaTyr..... 246
413 GAGGACAGAGTGTACACTGACCTGACCTGATCCGACGACGAC 462
247PheSerValGlyLysTyrPasnLeuSerGlySerAlaSerLeuLysG 262

463 GCCTCTCTCCGTCCTCCACAGCTCCCTGTGAGAGTGGGCTTGGCT 512
262 InGlnLysProPheLeuLeuIleSer.....ValIleGluGlu 275
513 TCGAGCAGCGCTACTTCATGTCAGTCAGTCCATGCTCTGTGTCATGGGA 562
276 GlyArg..... 277
563 GGCACCTTGGGAGACTTCTGCAACAGCGGCGCATGTCATGACACGTT 612
278 .MetIleSerGlyGluTyrValTyrProPheLysLysGlyAspHisMetL 294
613 CATAGCCAGCGGCGCAACCAAGTGGTGGCGCATGTCGATGT 652
294 euleuProTyr.....GlyLeuGly 300
663 TCGTGCACACAGCATWGGAGGCGCATGGGC 691
seq_name: gp_est88:BF488952

seq_documentation_block: 652 bp mRNA EST 06-DEC-2000
LOCUS BF488952
DEFINITION AT24629, Sprime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT24629 5 similar to CG11661:
Fban0011661 last updated: 000321, mRNA sequence.
ACCESSION Fban0011661
VERSION BF488952
KEYWORDS BF488952.1 GI:11572253
SOURCE EST.
ORGANISM fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephyridiidae; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
1 (bases 1 to 652)
Stapleton, M., Brokstein, P., Hong, L., Asbyani, A., Baxter, E., Berman
, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, D.,
Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Celniker, S.,
Lewis, S.E. and Rubin, G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
COMMENT
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AE003472; hit genomic sequence AE003525
Plate: AT.246 row: C column: 5
High quality sequence stop: 585.
Location/Qualifiers

FEATURES

source

1. 652
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT24629"
/lab_host="DH5-alpha or DH5-alpha Tona as per database (AT
121 on are in Tona cells)"
/note="Organ: Adult testes; Vector: pOTB7; Site_1: EcORI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 134 a 210 c 184 g 124 t
ORIGIN

alignment_scores:
Quality: 91.00 Length: 189

210 euLysSerIleGluVal.....IleGluValProSerIlePro 223
 423 TGAAGAAACATATTCACATTTACGCTTCATCA..... 466
 224 GUAUgHsThrValHsHsIsgIuInIleGluAspLeuLeuThrTh 240
 467ACAAATGTCCAAAGAAAGAAATCAACGGCTTAGTTGATGCAAC 507
 240 rLeuIleGlu.....CysAlaTyrPheSerValGlyLysT 252
 508 AATCATTAAGATGGCTCCATGATGCTGACATAGTTAATAGTAACT 557
 252 rPAsnLeuSerGlySerAlaSerLeuLysGlnIleLysProPheLeu 268
 558 GGCATTGGANAGTATTAAGTGGCGGAAGAAAGAAACTAGCATTCGA 607
 269 IleSerValIle 272
 608 CTGAGATCTC 619

seq_name: gb_est71:BE249341

seq_documentation_block: 531 bp mRNA EST 21-DEC-2000
 LOCUS BE249341
 DEFINITION NF014C07LF1053 Developing leaf Medicago truncatula cDNA clone
 NF014C07LF 5', mRNA sequence.

ACCESSION BE249341
 VERSION BE249341.2 GI:11959569
 KEYWORDS EST.
 SOURCE Medicago truncatula
 ORGANISM Medicago truncatula

REFERENCE
 AUTHORS Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
 Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids 1;
 Fabales: Fabaceae: Papilionoideae: Medicago.
 1 (bases 1 to 531)
 Flores-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
 Torres,H.R., Inman,J.T., Weller,J.W. and May,G.D.

TITLE
 JOURNAL Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula leaf library
 Unpublished (2000)
 On Jul 13, 2000 this sequence version replaced gi:9119401.

COMMENT
 Contact: May GD
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2310 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7391
 Fax: 580 221 7380
 Email: gdmay@noble.org
 Medicago Genome Initiative accession: MGI:S:14889
 Insert length: 659 Std Error: 0.00
 Plate: 014 row: C column: 07
 Seq primer: TCACACAGAAACGCTATGAC.

FEATURES
 Location/Qualifiers
 1..531
 source

/organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF014C07LF"
 /clone_lib="Developing leaf"
 /tissue_type="leaf"
 /dev_stage="Pooled developmental"
 /note="vector: lambda zap; Contains a mixture of very
 young, developing, mature and senescing leaves."

BASE COUNT 165 a 95 c 134 g 137 t

ORIGIN

alignment_scores:
 Quality: 89.00 Length: 146
 Ratio: 1.290 Gaps: 6
 Percent Similarity: 47.260 Percent Identity: 25.342

alignment_block:
 US-09-462-846-2 x BE249341 ..

Align seg 1/1 to: BE249341 from: 1 to: 531

80 ProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeuSerValGlnVal 96
 4 CCTCTCTTCAATCAATAAAGCTTGCCTAAACCCAGATCTCACCTCATCTGT 53
 96 LHisProAsnAspGluTyrAlaAsnIleHsIsgLusnGlyGlu..... 110
 54 TCAT...AACACCGTCGATAGAAATGATGATGAGAGAGGTAAGGCTGAAC 100
 110 110
 101 AGTGGTTGAAGTGTATCAATGAGGAGCTAGAGCTTTGTATCAAT 150
 111 ..LeuGlyLysThrGlyLysTyrTyrIleIleAspGlyGlnLys..... 124
 151 TTCTGACCAAGAGAGATGTGATGATGATGATGATGATGATGATGATGATG 200
 125AspAlaGluIleIleTyrGlnHisAsn 134
 201 TATGACACAGTCAACGGTGTGTATGACGACACTGGAAGCAAGACA 250
 134 IaThrThrLysGluGluLeuThrThrMerIleGluArgGlyGluTyrAsp 150
 251 GCAGAGTGTCTACTAGCTCTGGAACATTTCTGCGCCAGGGAGCT...GAT 297
 151 GluLeuLeuArgArgValLysValLysProGlyAspPhePheTyrValPr 167
 298 AAGATTGTGAGGATATTTGAAAAAAATTCGCCACTTACTTATTTATTC 347
 167 cSerGlyThrValHisAlaIleGlyLysGlyLeuAlaLeuGluThrG 184
 348 T.....GTAGACACCGGTGAAGACTTCAAGTCTACTTATGAAATG 391
 184 IngInAsnSerAspThrThrTyrArgLeuTyrAspTyr 196
 392 GACAAAGATAGAGCTCTAC.....TATGACTAC 420

seq_name: gb_est88:BF493329

seq_documentation_block: 712 bp mRNA EST 06-DEC-2000
 LOCUS BF493329
 DEFINITION AT01657.5prime AT Drosophila melanogaster adult testes pothB7
 Drosophila melanogaster cDNA clone AT01657.5 similar to CG11661:
 FBan0011661 last updated:000321, mRNA sequence.

ACCESSION BF493329
 VERSION BF493329.1 GI:11576630
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster

REFERENCE
 AUTHORS Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
 Pterygota: Neoptera: Endopterygota: Diptera: Brachycera:
 Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
 1 (bases 1 to 712)

Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
 B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
 D., Frisze,E., George,R., Gonzalez,M., Guarini,H., Harris,N., Li,P.,
 Iiao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Paclebo,J.,
 Park,S., Paragas,V., Phoumenavong,S., Wan,K., Yu,C., Celisner,S.,
 Lewis,S.E. and Rubin,G.M.
 Berkeley Drosophila Gene Collection Project
 Unpublished (2000)
 Contact: Stapleton, M.

TITLE
 JOURNAL BDGP
 COMMENT

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
 hit genomic sequence AB003472: hit genomic sequence AB003525
 Plate: AT.16 row: E column: 9
 High quality sequence stop: 479.
 Location/Qualifiers

FEATURES

```
source 1..712
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_1ib="AT01657"
        /dev_stage="0-3 day old Ore-R males"
        /lab_host="DHS-alpha or DHS-alpha Tona as per database (AT
121 on are in Tona cells)"
        /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 143 a 228 c 204 g 137 t
ORIGIN

alignment_scores:
    Quality: 88.50      Length: 202
    Ratio: 0.932      Gaps: 8
    Percent Similarity: 47.030      Percent Identity: 23.762

alignment_block:
US-09-462-846-2 x BF493329 ..

Align seg 1/1 to: BF493329 from: 1 to: 712

155 ArgValLysValLysPro..... 160
|||||:|||||:
86 CCGCTGAACACTGTCGCCCATGCATGACGACGACACCCCTGAAGACGAT 135
|||||:|||||:
161 GLVSPHPHETRYVALPROSERGLYTHRVALHSAIAIE.....G 175
|||||:|||||:
136 TGGCAACATGTTCTCGACTCTCCGCCCGCGGACACAGATCGAGACAC 185
|||||:|||||:
175 LYSGLYIIEUALALEUGLUHRLINGLINAASERAPYTHRYT 191
|||||:|||||:
186 ACNAGGCGATCTTCGCATCTGCGCCGCGCAGCATGTCGACAGAC 235
|||||:|||||:
192 ARGLEUTYRASPTRYASPARGLYSAPALA..... 201
|||||:|||||:
236 AAGTGGGGGACTGTCGCTGGGAGACCTTCGCTCGCTCGCTCCT 285
|||||:|||||:
202 ...GLUCILYLSLEUARGGLUENHISLEULYSERILEGLUVALI 217
|||||:|||||:
286 TAAGGAGGGC.....ATCCACGTCGACTATCCGGCCAGGATG 323
|||||:|||||:
217 LEGLUVALPROSERILEPROGLUARHISTHRYVALHISGLUINILE 233
|||||:|||||:
324 TGGACGCGGACCTTTTCGACGCGGACGACGCTCTGCACACCACTCC 373
|||||:|||||:
234 GLUASPLEULEUTHRTHRLEULIEGLUCYALATYR..... 246
|||||:|||||:
374 GAGGCAAGGTGTGTACAACTCAGTGGACCACTTTATCCGACCAAGGC 423
|||||:|||||:
247 ...PHESERVALGLYSTRPASLEUSERGLYSERIALASERLEULYS 262
|||||:|||||:
424 GCCCTACTCCGTCCTTCAACAGCTCCCTGCGAGTGGCTCTGGCT 473
|||||:|||||:
262 INGLILYPROPHLEULEULESER.....VALILEGLUGLYGLU 275
|||||:|||||:
474 TCGACACTGCTACTCCATGCGCCAGTCCCAATCTGTGTCATGTGGGA 523
|||||:|||||:
276 GLYATG..... 277
|||||:
524 GGCCAGTTTGGGACTTCTGCACACGCGCGCAGTCATCATGACACGTT 573
|||||:
278 METILESERGLYGLUTRYVALTYRPROPHETLSYGLYASPHISMETL 294
|||||:
574 CATACCCAGCGCGGACGACAGTGGTGGCGCCGAGTCGATCATGT 623

294 euleuProTYR.....GlyLeuGlyGluPhelLysLeuGluGlyTyr 307
|||||:|||||:
624 TGCTGCCACACAGCATGAGAGGCGATGGCCCGGACGACACTCTCCGCGCGC 673
|||||:
308 MlaGlu 309
|||||:
674 ATGAG 679
```

Fri May 18 10:38:05 2001

us-09-462-846-2.p2n.rst

Page 13

•
•
•
•


```

518 GCAGCAGCCCAATGCTGAATCTG...ATGGCGCGCATCCCAAGCA 564
47 lnservValValAlaAsnGlyMetTyrIsgLysPheThrLeuSerGlyLeu 63
   |||:::|||||:::
565 GCTCCGATCACCACCGCCAGCGCAACCGCTCTCCCTCCGTCGACGCC 614
64 TPTGluHisArg...HisLeuPheGlyGlnLeuGlyLysAspArgPhe 79
   |||:::|||||:::
615 ATCGAAAGAAATAAACCGCATCTGGCGAGAGCGGTAGCCACCTGT 664
79 e.....ProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeu 93
   |||:::|||||:::
665 CGCGCAACTGCGCTTCTGTTAAAGTACTGCGCGCCCAACCCCTCT 714
715 CTATTCAGGTGACCCGAAAT.....AAACCAACTCC 746
110 GluLeuGly..... 112
747 GAAATCGGTTTGGCAAAAGAAATGCGCGGCTATCCCATGATGCC 796
113 .....LysThrGlyCysTyrTyrIleI 120
797 AGAGCGGAACATTAAAGATCTTAACCATTAACAGAGCTGTTTGGCC 846
120 leasp.....CysGlnLysAspAlaGluIle 129
   :::::
847 TGACCGCTTCTGCGCATGAACCGCTTCCCGCAATTTCTGCATGTC 896
130 Tyr.....GlyHisAsnAla..... 134
897 TCTTACTGCACTGTGCGCGCGCATTCGCTATCGCCACTTTT 946
135 .....ThrThrLysGluLeuThrPheMetIle..... 144
   :::::
947 GCAGTGGCGAATGCTGAACCTGACGACGCTTTTCCAGCGCTTGA 996
144 ..... 144
997 AATGCAAGCGCAAGAAATCCGCGCGCTTACCGCTACTCAAGCGCG 1046
145 .....GluArgGlyGlu...TyrAspGlu..... 151
   :::::
1047 CTTAAGACGACAGAGCGCAACCTGGCAAAAGATCCGCTGATTGACA 1096
152 .....LeuLeuArgArgY 156
1097 GATTAATCCTGACGACAGCGGCTTTCTCTCTTGTGCTGATGATG 1146
156 allYsValLysProGlyAspPhePheTyrValProSerGlyThrValHis 172
   |||:::|||||:::
1147 TCAAACTGAATCCGCGAGCGCATGTTCTGTTGCTGAAAGCGCTCAT 1196
173 AlaIleGlyLysGlyIleLeuAlaLeuGlnThrGlnLysSerPhe 189
   |||:::|||||:::
1197 GCTTATCTGCAAGCGCTT...GCCGTGAAGTATGCGCAACTCGATTA 1243
189 rThrTyrArgLeu..... 193
   :::::
1244 CGTCTGCGCGCTGCTTACGCCAAATATATGACATCCCTGAGCTGG 1293
194 .....TyrAspTyrAspArgLysAspAlaGluGlyLysLeuArgGlu 207
1294 TGGCAAGCTGAAGTTGCAACCTTAAGCTGCG...GGCGGTGCTGACT 1340
208 LeuHisLeuLysLysSerIleGluVal...IleGluValProSerIlePhe 223
   :::::
1341 GCCCGGTGAAGAAAGCGCGCGAGCTGACTCCCAATTCGGGTGACGA 1390
223 oGluArgHisThrValHisHisGluGlnIleGluAspLeuLeuThrTh 240
   :::::
1391 TTTTCTTTTCACTGACGACGCTGCGCTTCCAGGAGCG..... 1430

```

```

240 hLeuIleGluCysAlaTyrPheSerValGlyLysTyrAsn..... 253
1431 .....AGCATCGGCCAACACAGCCCGCGATT 1457
254 .....LeuSerGlySerAlaSerLeuLysGlnGlnLysProPhe 267
1458 CTGTTCTCGTTGAGCGGTGAGCGCGGTGTACTGAAGATGAACAGCGCT 1507
267 uLeuLeuSerValIleGluGlyGluGlyArgMetIleSerGlyGlu 282
1508 GGTACTGAAG.....CCGGTGAATCTGCTTATCGCGCGCGAT 1547
seq_name: /cgn2.6/prodata/1/lna/6R.COMB.seq:US-09-001-982-9
seq_documentation_block:
; Sequence 9, Application US/09001982
; Patent No. 6204246
; GENERAL INFORMATION:
; APPLICANT: Bosch, Hendrick J.
; APPLICANT: Stiekema, Willem J.
; TITLE OF INVENTION: Hybrid Toxin
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6204246artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,982
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/602,737
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 130-4080/PCT/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8687
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3468 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3468
; US-09-001-982-9
alignment_scores:
Quality: 92.50 Length: 338
Ratio: 0.571 Gaps: 18
Percent Similarity: 47.929 Percent Identity: 18.639
alignment_block:
US-09-462-846-2 x US-09-001-982-9
Align seg 1/1 to: US-09-001-982-9 from: 1 to: 3468

```



```

598 GACGAGGCTCCAAAT...GACATCATGAAATGTCAGGATGATCATG 644
34 hcglyglucystrpalaphleahlaahsglnsnglyglnserValVal 50
   |||||
645 CGGAGAAATGCTCATCCAAAGCAACGATAT..... 673
51 GlnasnGlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGlnHisH 67
   |||||
674 .....CATAAAAATTCATGTTTTCGAGCTTACGACCAATCA 711
67 sarghHisLeuPheGlyGlnLeuGlnGlyAspArgPheProLeuThrL 84
   |||
712 TGA.....CGAGAAAAAATTCCTCCCTGCTCCAT 743
84 ysIleLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAsp 100.
   |||||
744 CACTGTTAAGATCA.....ACTGTCTCAAGCGTTTCATACAACTAAC 787
101 GlnTyrAlaAsnIleHisGlnAsnGlyGluLeuGlyLysThrGluCysTr 117
   |||||
788 ...TTTACTAAGCCACATCAAC..... 808
117 PTrIleIleAspCysGlnLysAspAlaGlnIleIleTyrGlyHisAsnA 134
   |||||
809 .....CTGGATACCGAA 821
134 IaThrThrLysGluGlnLeuThrThrMetIleGluArgGlyLutPasp 150
   |||||
822 CGTCTACATCGGTCCACTGTGTGTG..... 847
151 GlnLeuLeuArgArgValLysValLysPheGlyAspPhePheTyrValPr 167
   |||||
848 GAATATCTACAGCGTAGATCTGTATACCGTATGATCTTGCATGCAATGCG 897
167 oSerGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrG 184
   |||||
898 GACAGGTATACAGTAGAAATTTCTCTTTTACCAAAAACAGCAGCG 947
184 IngIlnAsnSerAspThrThrTyrArgLeuTyr..... 194
   |||
948 GACCAAGCGCTCACAGTCTACAGAGATAGATTTCGAATCGCA 997
195 AspTyrAspArgLysAspAlaGlu...GlyLysLeuArgLys..... 207
   |||||
998 AATTATCAAGTCAGGATTGGAAACCGAATAATAGACCCCTAAAAA 1047
207 ..... 207
1048 AAGAAACTTCTAACAGATGAACATTCACATAGCGTGGATGCAATGG 1097
208 .....LeuHisLeuLysSerIleGluValIleGlu 218
   |||||
1098 AAGAAAGAGATCTGTATGATCTCTCAGTAAAGATGAGATGAGCCGGAA 1147
219 ValProSerIleProGluArgHisThrValHisIleGlnIleGluAs 235
   |||||
1148 GCAGTTCGTTCGTACAAAAACAGT...TACCACTTTCACCTTAAGA 1194
235 PleuLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLys. 251
   |||||
1195 TATGACTGTGACG.....TTCTGCTCCGGAAGAAC 1223
252 .....TrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPhe 266
   |||||
1224 AACCTTTTAACATCAGC...AGCTCATTTGGCTGAATGCGTCTTACC 1270
267 LeuLeuIleSerValIleGluGly..... 274
   |||||
1271 ATAGCCACGAGGCGCATAGATGCGATCTTTCAGAAAGATAGATTGCGAC 1320
275 .....GlnGlyArgMetIleSerGlyGluTyrV 284
   |||||
1321 TCATGTCCGTTCTGTGGGACATCGAAATACATCTCGAGATGGGCGATTTC 1370

```

```

284 aLlyrProPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGly 300
   |||||
1371 TCATGCCATTTTCAGAA.....CTCATGACCATGCTTGCTTGGCT 1408
301 GluPheLysLeuGlu 305
   |||||
1409 GAATGCTACCTAGAA 1423

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-813-940-24

seq_documentation_block:

```

; Sequence 24, Application US/08813940
; Patent No. 5834279
; GENERAL INFORMATION:
; APPLICANT: Rubin, Harvey
; APPLICANT: Yang, Fude
; APPLICANT: Avarbuck, David
; APPLICANT: Curran, Sean
; TITLE OF INVENTION: Methods of Identifying Compounds that
; TITLE OF INVENTION: Inhibit DNA Synthesis in Mycobacterium Tuberculosis and
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5834279-15
; STREET: One Liberty Place, 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,940
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,271
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPAP-0220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 446..2620
; US-08-813-940-24

```

alignment_scores:

```

Quality: 84.00 Length: 276
Ratio: 0.672 Gaps: 12
Percent Similarity: 45.280 Percent Identity: 22.826

```

alignment_block:

US-09-462-846-2 x US-08-813-940-24 ..

Align seg 1/1 to: US-08-813-940-24 from: 1 to: 4107

```

GENERAL INFORMATION:
APPLICANT: Schlössinger, Joseph
APPLICANT: Yan, Hai
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PROTEIN
TITLE OF INVENTION: PHOSPHOTYROSINE PHOSPHATASE-SIGMA
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,464
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/130,570
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-043
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5690 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 833..5338
US-08-447-464-2

```

```

alignment_scores:
    Quality: 84.00    Length: 211
    Ratio: 0.778      Gaps: 11
    Percent Similarity: 51.185    Percent Identity: 21.327

alignment_block:
US-09-462-846-2 x US-08-447-464-2 ..

Align seg 1/1 to: US-08-447-464-2 from: 1 to: 5690

40 pheAlaIahIsGlnAsnGlyGlnSerValValGlnAsnGlyMetTyrLys 56
   |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
4091 TTTTCTCTACACAGATGGCTGTAGTACAGAAAGGTGAGTACGACATTT 4140
   |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
56 sGlyPheThrLeuSerGluTyrLysTrpGlnHisAlaGlnHisLeuPheGly 73
   |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
4141 TCACTTACAA.....GCATGGCTGACCAC.....GGGG 4165
   |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
73 ILeuGlnGlyAspArgPheProLeuLeuThrLysIleLeuAspAlaAsp 89
   |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
4170 TACCCGGAG.....TACCC.....ACACCTTCTCTGGCGTTGTG 4204
   |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||::
90 GlnAspSerValGlnValHisProAsnAspGluTyrAlaAsnIleHis 106
   ::::: ::::: ::::: |||::: |||::: |||::: |||:::
4205 CGCAGAGTCAGACACTGCAACCCGCGCTACCGCTGGCCACGATTGGTGCCA 4255
   ::::: ::::: ::::: |||::: |||::: |||::: |||:::
106 sGlnAsnGlyLeuLeuGlnGlyLysThrGlnCysTrpTyrIleLeuAspCys 122

```

```

      | ..... :|||::||| ||::||| .....  

4255 GTGACCGCGGGGTGGGGGGCTACTGAGTCTTCATTTGAATTGATGCCA 4300  

123 ..... Glnlyspaslaelutlleterytqlyhis 132  

4305 TETTGAGCGCATCAGAACAGAGAAGCGGTGAT...GTGACGAGACC 4351  

133 AsnAlarThrLysglugluLeuThrMetileGuargglyGluTr 149  

      :::::::::::::::::::: :|||::||| .....  

4352 GTGACACTCATCAGCGGTGCAGACCCACTACATGATGTCAGACGAGAT... 4399  

149 pAspIleuLeuArgArGylValLysValLysProGlyAspPhePheTyv 166  

4400 ..... CgGTRAcCTTCA 4412  

166 alProSerGlyThrValHisAlaIleGlyLysGlyIleLeuaLaenglu 182  

4413 TCACAGAGGCACTCTGTGAGGCTGTGGCGTGTGC..... 4444  

183 ThrGInGaAsenSerAPThrThrTyAr...LeuTyArAsPTyrAsPaR 198  

4448 .....ATPACCGAGGTCCCCGCCGCGAGGCTGTACAACTTAATCCA 4488  

198 gLyAspaAlaGlu.GlyLysLeuArgGluLeuHisLeuLysLysSerile 214  

4489 GAAGCTGGCCCGAGCTGAGGCTGGCGAGCATGTCAAGAAAGGAGCTTG 4538  

215 GluValIleGluValProSerIleProGluArgHisThrValHisHISgl 231  

      :::::::::::::::::::: :|||::||| .....  

4533 AGTTACAGAGGCTTGCCAGCT.....CCAAGCACACACTTCGAGATTTC 4582  

231 uGlnIleGluAspLeuLeuThrThrTrHeu 241  

      ::::::: ||::|||::||| .....  

4583 ATCACTGCCACCTCTGCTGCACAACAAGTTTA 4613  

seq_name: /cgr2_6/prodat4/1/ina/5B_COMP.seq;US-08-716-679-2  

seq_documentation block:  

Sequence 2, Application US/08716679  

Patent No. 5840800  

GENERAL INFORMATION:  

APPLICANT: Schlessleringer, Joseph  

APPLICANT: Yan, Hai  

TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PROTEIN  

TITLE OF INVENTION: PHOSPHOTYROSINE PHOSPHATASE-SIGMA  

NUMBER OF SEQUENCES: 12  

CORRESPONDENCE ADDRESS:  

ADDRESSER: Pennte & Edmonds  

STREET: 1155 Avenue of the Americas  

CITY: New York  

STATE: New York  

COUNTRY: U.S.A.  

ZIP: 10036-2711  

COMPUTER READABLE FORM:  

MEDIUM TYPE: Floppy disk  

COMPUTER: IBM PC compatible  

OPERATING SYSTEM: PC-DOS/MS-DOS  

SOFTWARE: PatentIn Release #1.0, Version #1.25  

CURRENT APPLICATION DATA:  

APPLICATION NUMBER: US/08/716,679  

FILING DATE:  

CLASSIFICATION: 435  

PRIOR APPLICATION DATA:  

APPLICATION NUMBER: US/08/130,570  

FILING DATE:  

ATTORNEY/AGENT INFORMATION:  

NAME: Mastrock, S. Leslie  

REGISTRATION NUMBER: 18,872  

REFERENCE/DOCKET NUMBER: 7683-043  

TELECOMMUNICATION INFORMATION:  

TELEPHONE: 212-790-9990  

TELEFAX: 212-869-8864/9741  

TELEX: 66141 PENNTE

```

```

? ? ? ? ? INFORMATION FOR SEQ ID NO: 2:
? ? ? ? ? SEQUENCE CHARACTERISTICS:
? ? ? ? ? LENGTH: 5690 base pairs
? ? ? ? ? type: nucleic acid
? ? ? ? ? STRANDEDNESS: unknown
? ? ? ? ? TOPOLOGY: unknown
? ? ? ? ? MOLECULE TYPE: DNA (genomic)
? ? ? ? ? FEATURE:
? ? ? ? ? NAME/KEY: CDS
? ? ? ? ? LOCATION: 833..5338
? ? ? ? ? US-08-716-679-2

alignment_scores:
? ? ? ? ? Quality: 84.00 Length: 211
? ? ? ? ? Ratio: 0.778 Gaps: 11
? ? ? ? ? Percent similarity: 51.185 Percent identity: 21.327

alignment_block:
US-09-462-846-2 x US-08-716-679-2 ..

Align seg 1/1 to: US-08-716-679-2 from: 1 to: 5690

40 PhealaalaahisglnaenlglylnserivaigaInasnGlymetrily 56
||||| :|||||:|||||:||||| :||| :||
4091 TTTTCTACACAGAAATGGCTCTAGAGAGAAACCGAGGTACGACATT 4140
56 sGlypherthleuseretoleutrpGlnHshIstahIstleupneIyG 73
||||| ||| ||| |||
4141 TCAGTTCCCA.....GCATGGCGTCGACGC.....GGGG 4169
73 InleuGlucIyAspArgPheProleuLeuthrIySileuAspAlaasp 89
||| ||| ||| |||
4170 TACCCGAG.....TACCC.....ACACCTTCCTCGCGTTTCG 4204
90 GlnAspSerVaIglInVaIHisProAsnAspGluTyrAlaAsnIleH 106
::: |||:|||||:|||||:|||||:|||||:|||||
4205 GCAGAGTCGACAGCTCGAACCCGCTGACGCTGGCCGCGCTTGCTGCA 4254
106 sGluAsnGlyIuLeuGlyIystrGluCystrPyrIleIleAspCyS 122
| :|||:|||||:|||||:|||||:|||||:|||||
4255 CTGCAGCGCGGGGTGGGGCGGTACTGGCTGCATTGTAATTGATGCCA 4304
123 .....GlnIyAspAlaIuIleIytrGlnHs 132
|||:|||||:|||||:|||||:|||||:|||||
4305 TGTTCGAGCGCATAGACAGACAGAAAGCGTGGAT..GTGACGGACAC 4351
133 AsnAlaThrThrIySgluIuLeuThrThrMetIleGluArgIyCyluTr 149
::: |||:|||||:|||||:|||||:|||||:|||||
4352 GTGACACATCATCGCGTACAGCGCCACTACATCATGTGTCACAGAGATT 4399
149 pasProIuLeuIuArgIyValIySValIySProGlyAspPhePheTyr 166
.....CAGTTCACCTCA 4412
4400 .....CAGTTCACCTCA 4412
166 alProSerGlyThrValHisAlaIleGlyIySgluIleuAlaIeuglu 182
::: |||:|||||:|||||:|||||:|||||:|||||
4413 TCCAGAGGCGCTCTGTGAGCGTGTGGGGCTGTGC..... 4447
183 ThrGlnGlnAsnSerAspThrThrTyrTyr..LeuTyrAspTyrAsp 198
||||| ||| ||| ||| ||| ||| ||| |||
4448 .....AATACCGAGGTCCCGCGCGCGCAGCCTTACACCTAATATCA 4488
198 gLyAspAlaGlu.gLyIySleuArgIuLeuHshIleuIySlysSerIle 214
||| |||:|||||:|||||:|||||:|||||:|||||
4489 GAACGTGGCCAGGTGGAGGCTGCGACAGATGTCACAGAAATGAGAGCTTG 4538
215 GluValIleGluValProSerIleProGluArgHshIstValHisHsGI 231
::: |||:|||||:|||||:|||||:|||||:|||||
4539 AGTTCAGAGGCTTGCCAGCT.....CCAAAGCCACACCTTCGAGATTG 4582
231 uGlnIleGluAspLeuLeuThrThrIleu 241
::: |||:|||||:|||||:|||||:|||||:|||||

```



```

8346 .....CTGTTCCACCCAGACGAGACGAGCTGTTCAAGA 8307
281 lylglutylvaltyrprophelyslyslvaspishmetleuprotyr 297
8306 GAGAAATCGAAGACGTAGTGGACAGAGACGCGCTGAGACCCGTT 8257
298 glyleuglygluphelysleu...gluglytyrvalaglu.....cy 310
8256 GGTCTCGCCGGAAGTATCGCTCGATGAGGAAATGCGAGCTGAGACTACGC 8207
310 sllevalserhis 314
8206 TCTGCTCTCCAC 8194

seq.name: /c962_6/prodata/1/lna/5B.COMB.seq:US-08-222-617A-24
seq.documentation block:
: Sequence 24, Application US/08222617A
: Patent No. 5882879
: GENERAL INFORMATION:
: APPLICANT: Veenstra, Annemarie E.
: APPLICANT: Martin, Juan F.
: APPLICANT: Garcia, Bruno D.
: APPLICANT: Gutierrez, Santiago
: APPLICANT: Barredo, Jose L.
: APPLICANT: Von Doehren, Hans
: APPLICANT: Palisa, Harriet
: APPLICANT: Van Liempt, Henk
: TITLE OF INVENTION: A Method for Influencing Beta-Lactam
: TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
: STREET: 300 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/222,617A
: FILING DATE: 04-APR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: REFERENCE/DOCKET NUMBER: 97,157
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 11601 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Acremonium chrysogenum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 388..11526
: OTHER INFORMATION: /function= "Enzyme"
: OTHER INFORMATION: /product= "ACV Synthetase"
: NAME/KEY: CDS
: LOCATION: 8050..8052
: OTHER INFORMATION:
: OTHER INFORMATION: /note= "NNN-AGU, AGC, UCU, UCC, UCA, or UCG"
: US-08-222-617A-24

```

```

alignment_scores:
  Quality: 84.00      Length: 355
  Ratio: 0.503      Gaps: 25
  Percent Similarity: 47.042      Percent Identity: 23.380

alignment_block:
US-09-462-846-2 x US-08-222-617A-24/rev ..

Align seg 1/1 to reverse of: US-08-222-617A-24 from: 1 to: 11601

34 ThrGlyGluCysTTPAlaPheAlaHisGlnAsnGlyInSerValVa 50
9133 ACGGGGGAAGGTGGCGTGCAGAGGTGCAC.....GGATTCGATGCT 9090
50 lGlnasnGlyMetTyrLysGlyPheThrLeuSerGluLeuTyr..... 64
9089 TCGCACACAGCTCCGGCGCTGAGCGTGTAGTACGACGACATGGCGCTGAC 9040
65 .....GluHisHisArgHisLeuPheGly..... 72
9039 CCTGCTGATCGAGCCATCTCTCATCTTGGGATGACGCGACGCT 8990
73 .....GlnLeuGluGly...AspArgPheProLeuThrLysIleL 86
8989 CCTTAACAGACAGATGCTGACGAGCGGCACTCCGACCTCGAGAG..C 8943
86 euAspAlaAspGln.....AspLeu..... 92
8942 TCGATGCGGAACCCCTCAGCTTACCTGCTGTCTCGCGCGAGATA 8893
93 .....SerValGlnValHisProAsnAspGluTyrAlaAs 104
8892 CTCGAGATGTGGGTGCCACGAGAGCCAC.....CAGATTCACG 8852
104 nIleHisGlnAsnGlyGluLeuGlyLysThr...GluCysTTPtyrIleI 120
8851 TCTGTAGAGTCT.....CTGGGACAGCTGTCACTGCTGTCTTCCGC 8808
120 lAspCysGlnLysAspAlaGluIleIleTyrGlnHisAsn..... 133
8807 TCGTAGAAAGGTTGGGATGATGATCCCTGCTGTCAGGCACTCGTGT 8758
134 .....AlaThrThrLysGlnGluLeuThrThrIle 143
8757 GAGGTACCCGCGAGACGACGATCACGCCCAAGTAGAGTCCCAACAG 8708
143 tIleGluArgGlyLutyrPaspGluLeuLeuArgArgValLysValLysP 160
8707 CGTTGAAGAAACACGCTGAGTGTGCTTCAAGAGACGTGACTT 8658
160 roGlyAspPhePheThrValProSerGlyThrValHisAlaIleGlyLys 176
8657 CCGGGGAT.....CCCGTCGACAGTCCCTGGTAA 8626
177 GlyIleLeuAlaLeuGluThrGlnAsnSerAspThrThrTyr...Ar 192
8625 GGGGCGATCGCCCTTGAA.....CTGTGATGATGTTGTACACGCTGC 8582
192 gLeuTyrAspTyrAspArgLysAspAlaGluGlyLysLeu.....A 206
8581 TCTC.....AGTGATACCATAGCGCTTGTGATGTCGCC 8547
206 rGluLeu.....HisLeuLysLys.....SerIleGluVal..... 216
8546 GCGAACTGGAGCGCACTTCTCAAACTGATGACGTGGAACCTCCGCGC 8497
217 .....IleGluValProSerIleProGluArgHisThrVa 228
8496 CGCAAGCGTGACCAATGTGAAGATCGGACAGACGAGAGACTCAATCTGT 8447
228 lHisHis..... 231
8446 GCAGCACCGAGGCGTCCGCTGATGATAGTATGATCTTCCCTCGCGCGC 8397

```



```

1      TYPE: nucleic acid
2      STRANDEDNESS: double
3      TOPOLOGY: linear
4      MOLECULE TYPE: genomic DNA
5      US-08-902-632-3
6
7      Alignment_scores:
8          Quality: 83.00      Length: 355
9          Ratio: 0.532      Gaps: 21
10         Percent Similarity: 43.944      Percent Identity: 21.127
11
12         Alignment_block:
13             US-09-462-846-2 x US-08-902-632-3 ..
14
15         Align seg 1/1 to: US-08-902-632-3 from: 1 to: 2325
16
17         116  ILETRPGIUGLYTHRALALEUAlAspPheGLYTYRThIleProSer.. 31
18             ::::: ::::: ::::: ::::: :::::
19         1186  TTGTGGGAGAACATAGTGTACCTAGATTTTACGTCCTGTAACCTCTCAT 1235
20             ::::: ::::: ::::: ::::: :::::
21         32  .....GlnArgThrGlyGluC 37
22             ::::: ::::: ::::: :::::
23         1236  CATCATCACCCACACGCTGTGCGCGATACGCTCAACAGAGAGA...T 1282
24             ::::: ::::: ::::: :::::
25         37  YSTRPALAPhe..AlAlaNIISGLAsnGlyIAsnSerValValGlnAsn 52
26             ::::: ::::: ::::: :::::
27         1283  GCAAGAAATATGACGCTTGCCCCACAGAGTGCGCCACCGCTTCGCAAGAC 1332
28             ::::: ::::: ::::: :::::
29         53  GlyMetYrLYsglyPhe.....ThrLeuSerGluLeuTrpGlnI 66
30             ::::: ::::: ::::: :::::
31         1333  .....TTCCGAGATGATTCGCGAGCGCTGTTGGAGACCTCTTGAGAGA 1376
32             ::::: ::::: ::::: :::::
33         66  SHIsArGhIsLeuPheGlyGlnLeuGlnGlyAspArgPheProLeuLeuT 83
34             ::::: ::::: ::::: :::::
35         1377  GAGCGAGAGATTAAGAAGATGAAGCGCAGATGACCCGATGACGAG 1426
36             ::::: ::::: ::::: :::::
37         83  hLYsTIELeuAspAlAspGlnAspLeuSerValGlnValHisProAsn 99
38             ::::: ::::: ::::: :::::
39         1427  GGAACTCTCTGATTACAGCGAG...AGGCCATCAAGATCCTGGCAAC 1473
40             ::::: ::::: ::::: :::::
41         100  ASPGIUTYrAlAsnIIEHISGLAsnGlyGluLeuGlyLYsThrGlyC 116
42             ::::: ::::: ::::: :::::
43         1474  AGCTACTTC.....GATTACTACGGCTATGCAAGGCG 1505
44             ::::: ::::: ::::: :::::
45         116  S...TrpTYrIIEAspCysGlnLYsAspAlGluIIErGlyGln 132
46             ::::: ::::: ::::: :::::
47         1506  GCGCTGTGATC.....TGCAAGGAGTGTGCGAGACGGA 1539
48             ::::: ::::: ::::: :::::
49         132  IASuAlIThThrLYsgluGluLeuThrMetIIEgluTrpGlyGlu 148
50             ::::: ::::: ::::: :::::
51         1540  ..ACGGCTGGGAGAGGAGATACATACGATGCATCAAG... 1578
52             ::::: ::::: ::::: :::::
53         149  TRPAspGluLeuLeuArgValLYsValLYs..... 159
54             ::::: ::::: ::::: :::::
55         1579  .....GAGATPAGAGAAAGATACGCTTAAAGTATCTACAGCGAC 1622
56             ::::: ::::: ::::: :::::
57         160  .ProGlyAspPheTYrValProSerGlyThrValHisAlaIIEglY 176
58             ::::: ::::: ::::: :::::
59         1623  CGAGCGATTTTGTGCACAAATACCTGAGCGCATGCTGAACCGTCANAA 1672
60             ::::: ::::: ::::: :::::
61         176  yS.....GlyIIE 178
62             ::::: ::::: ::::: :::::
63         1673  AGAAGGCTATGAGTCTCTCAATATCAAGCGCAAACTCCGGGCGG 1722
64             ::::: ::::: ::::: :::::
65         179  LeuAlaLeuGluThrGln..... 184
66             ::::: ::::: ::::: :::::
67         1723  CTTGAGCTCGAGATACGAGGCTTCTACAACGCGGCTTCTGCACGAA 1772
68             ::::: ::::: ::::: :::::
69         185  .....GlnAsnSerAspThrTYrArgLeuT 194
70             ::::: ::::: ::::: :::::
71         1773  GAAGAAGTATGCGGTATAGACGAGGAAGCAAGATTAACAACGCGCGGAC 1822

```



```

3213 GGTCTGATACATCTTCGTGTCACAGCGTAC..... 3245
303 sleuGluclylTyralaGu...CysileValSerHis 314
|||||.....|
3246 .AAGAGGCGCTACGAGAGAGATCGTACGATCCAT 3281
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-961-083-105

seq_documentation_block:
; Sequence 105, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Chol et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 105:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-105

alignment_scores:
Quality: 81.00 Length: 341
Ratio: 0.570 Gaps: 11
Percent Similarity: 41.642 Percent Identity: 17.889

alignment_block:
US-09-462-846-2 x US-08-961-083-105 ..
Align seg 1/1 to: US-08-961-083-105 from: 1 to: 1879

26 glyTyrThrIleProSerGlnArgThrGlyLcylucStrpAlaPheAlaI 42
|||.....|
23 GGTCAAGCTTACCTGAAGACATCGGAAACGAAAGAGGCTGACTTATC 72
|||||.....|
42 aHisGlnsnglyGlnSerValaIaGlnsnglyMetTyrGlyGlyPheT 59
|||||.....|
73 AGAAAGACGAGAGACACCGTTCTCACTCAAGCAAACTGAGAGCGTTA 122
|||||.....|
59 hr..... 59
||
123 CTGAATAACGATTCACCTCCGACACCTACAGAAAGAACTGAAGTAGAC 172
|||||.....|
60 .....LeuSerGluLeuTrpGluHisHisArgHis 69

```

```

173 GAGGAACAAAGCCCTTCTAGCTGGATACACTTTTGAAGAAAGATGAA... 220
|||...|
69 sleuPheGlyLcyluGluValAspArgPheProLeuLeuThrLysIleL 86
|||||.....|
221 .....CAGCTCAAAAATAATCCAGAGCTACACGATGATGCT 254
||
86 eu.....AspAlaAspGlnAspLeuSer 93
|||...|
255 TAAAGAAACTGTAGATACAGTCGATGTCGATGCGACACAGCAAGCTCCA 304
|||...|
94 ValGlnValHisProAsnAspGluTyrAlaAsnIleHisGlnAsnGlyL 110
|||||.....|
305 GCAAGAACTACTCTCGAACAAGTAAAGGTGAGTGAAGAAACAAACAAA 354
|||||.....|
110 uLeuGlyLysThrGlnCysTrpTyrIleLeaSpCysGlnLysAspAlaG 127
|||...|
355 A.....GACAGCATGATGTTG 371
|||...|
127 LuIleIleTyrGlyHisAsnAlaThrThrLysGluLcyluLeuThrThrMet 143
|||||.....|
372 CTGCTGCTTATCTTGAAGAAAGCTGAAGGAAAGCTCTTCACTGCGGGT 421
|||...|
144 Ile.....GluArgGlyGluTrpAspGluLeuLeu 154
|||...|
422 GTAAACCAAGTAATTCCTTATGAACATTCGCGTGATGATGATGTTAAC 471
|||...|
154 GArgValLysValLysProGlyAsp..... 162
|||...|
472 TCGCTATTAATAAGCTTCGGATATGCTCCTTGCTGACAAATGATA 521
|||...|
162 ..... 162
522 CTGCTAAATAATCCTGCTTACCTCCTCTTGAAGATTAAACAAAGGAAA 571
|||...|
163 PhePheTyrValProSerGlyThrValHisAlaIleGlyLysGlyIleLe 179
|||||.....|
572 TACTTCAATGAGTAAGTAAATGCGCAATACCTGTGGTAACAAAGGTCA 621
|||||.....|
179 uAlaLeu.....GluThrGlnGlnAsnSerAspThrThrTyrArg.... 192
|||||.....|
622 AGCTTATATGATCAACTGCGCGTATATGTTACTCAAACTATAAAGCTA 671
|||||.....|
193 ..LeuTyrAspTyrAspArgLysAspAlaGluLcylLysLeuArgGluLeu 208
|||||.....|
672 CTGTTAAGTTTACGAAATTAAGACGCTGAAGCTGACTGACTAATCTA 721
|||||.....|
209 HisLeuLysSerIleGluValIleGluValProSerIleProGluArg 225
|||||.....|
722 GTTGTACTATAAATGTAGAC...ATCAACATCAATGATTAATCTCTAA 768
|||||.....|
225 GHisThrValHisHisGluGlnIleGluAspLeuLeuThrThrLeuI 242
|||||.....|
769 AGAAGACGTTCAAAAAGCCGTTGCGAGCAACGTTAAAGACAGTACGAG 818
|||||.....|
242 LeuGlyAlaTyrPheSerValGlyLysTrpAsnLeuSerGlySerAla 258
|||||.....|
819 TTCCAGACAGCTTACTAGAAAAGCCAG..... 847
|||||.....|
259 SerLeuLysGlnLysProPheLeuLeuIleSerValIleGluGlyL 275
|||||.....|
848 .....GGTGA 852
|||||.....|
275 uGlyArgMetIleSerGlyGluTyrValTyrProPheLysLysGlyAspH 292
|||||.....|
853 AGGT.....CCATTACAGCAGGATGTA 875
|||||.....|
292 IsMetLeuLeuProTyrGlyLeu 299
||
876 ACCATGTGATTCATACGAACTC 898
||
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-040-751-4

```

```

seq.documentation_block:
: Sequence 4, Application US/08040751
: Patent No. 5407825
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel M
: APPLICANT: Sick, August J
: TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
: TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 54078
: TITLE OF INVENTION: Lepidopteran-active Toxins
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DAVID R. SALIWANCHIK
: STREET: 2421 N.W. 41st STREET, SUITE A-1
: CITY: GAINESVILLE
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/040,751
: FILING DATE: 19930329
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: SALIWANCHIK, DAVID R.
: REFERENCE/DOCKET NUMBER: MA39.C1.D3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3522 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus thuringiensis
: STRAIN: aizawai
: INDIVIDUAL ISOLATE: PS81A2
: IMMEDIATE SOURCE:
: LIBRARY: lambdagem - 11 (tm) Library of August Sick
: CLONE: 81A2
: US-08-040-751-4

alignment_scores:
: Quality: 81.00 Length: 347
: Ratio: 0.529 Gaps: 19
: Percent Similarity: 44.092 Percent Identity: 19.885

alignment_block:
: US-09-462-846-2 x US-08-040-751-4
:
Align seg 1/1 to: US-08-040-751-4 from: 1 to: 3522
:
54 MetTyrLysGlyPheThrLeuSerLysLeuTrpGluHisHisArgHisLeu 70
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2173 TGTATCAAAAATAGATGAGTCGAAATTAAGCCATTAAACCGTTAC.. 2220
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
70 urPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrLysIleLeu 87
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2221 .....CAATTAAGAGG.....TATATCG 2239
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
87 sPaLaSpGlnAspLeuSer.....ValGlnValHisProAsnAsp 100
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2240 AAGATGTCAGACCTTAGAAATCTATTATTTCGTCACAAATGCAAAACG 2289

```

```

101 GluTyrAlaAsnIleHisGluAsnGlyGluLeuGlyLysThrGlyCysTr 117
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2290 GAAACAGTAATGTATCCAGGTACGGGTCTTA.....TG 2324
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
117 PTrIle..... 119
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2325 GCCGCTTCAGTCGAAAGTCCAAATTTGGAGGTGGGAAACCGAATGGT 2374
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
120 .....leAspCysGln..LysAsp 125
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2375 GTGTCCACACCTTGATGATGAATCCTGATTAGATTGTTCTCGACAGAG 2424
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
126 AlaGluIleIleTyrGlyHisAsn..... 133
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2425 GGGGAAAATGTGCACATCATTCCTCATTTCTCCTTGACATTGATGT 2474
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
134 ....AlaThrThrLysGluGluLeuThrThrMetIleGluArgGlyLut 149
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2475 TGGATGCACAGACTGCAAGAG.....GATCTAGGCGTGT 2509
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
149 rPaSpGluLeuLeuArgArgValLysValLysProGlyAspPhePheTyr 165
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2510 GG.....CTGTATTCAAGATTAGACGACGAGAGGT.....TAT 2544
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
166 ValProSerGlyThrValHis.....AlaIleGlyLys 176
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2545 GCAAGATTAGAAATCTGGAATTTATCGAAGAAACCATTTATTTGAGA 2594
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
176 sGlyIle..... 178
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2595 AGCAGTGTCTGTCGTAAGACGCGGAAAAAATGAGACAAACGGG 2644
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
179 ....LeuAlaLeuGluThr.....GlnGlnAsn 186
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2645 AAAAAGTCAATTTGAAACAAACGAGTATATACAGAGGCAAAAGAGCT 2694
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
187 SerAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspAlaGlu 203
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2695 GTGATGCTTATTATTCGATATCTCAATATGATAGTTTCAAAGCAGATAC 2744
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
203 LysLeu.....ArgGluLeuHisLeuLysLys 213
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2745 AAACATTTGATGATTCATCGCGCAGATAGACTTGTTCATCAGATCCAG 2794
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
213 erIleGluValIleGluValProSerIleProGluArgHisThrValHis 229
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2795 AGGCTTATCTCCAGAACTACCTTTCATTCACAGGAATAATGTGGTATT 2844
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
230 HisGluGlnIleGluAspLeuLeuThrThrThrLeu..... 241
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2845 TTGGAAGAAATTAGAAACCGTATTTCACGCAATATTCCTATATGATGC 2894
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
242 .....IleGluCysAlaTyrPheSerValGlyLys.....TyrA 253
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2895 GAGAAATGTATTAATAATGCGCATTTCAATTAATGCTATATCATGCTGA 2944
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
253 snLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeu 269
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2945 ACGTGAAGGCGCATGTAGATGTAGTAGAACAAACAAACACCGTGGTGC 2994
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
270 SerValIle...GluGlyGluGlyArgMetIleSerGlyLysTyrValT 285
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2995 CTGTTGTCCCGGAATGGAAGGAGAGGTGCACAACAAATTCGTGCTGTG 3044
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
285 rProPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyLup 302
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
3045 TCCGCGCGGTGCTATATCTCCGCTGTTACAGCGTAC..... 3081
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
302 helYsLeuGluGlyTyrAlaGlu...CysIleValSerHis 314
: ::::::::::: ::::::::::: ::::::::::: :::::::::::
3082 .....AAGAGGATATGAGAGAGGTTGCTTAACCATTCAT 3117

```

seq_name: /cgn2_6/prodata/1/lna/5A.COMB.seq:us-08-291-368-1

seq_documentation block:

Sequence 1, Application US/08291368

Patent No. 5686069

GENERAL INFORMATION:

APPLICANT: Payne, Jewel M.

APPLICANT: Sick, August J.

TITLE OF INVENTION: No. 5686069el Bacillus thuringiensis Isolates

TITLE OF INVENTION: Active Against Lepidopteran Pests

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sallwanchik & Sallwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: US

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/291,368

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/597,607

FILING DATE: 15-OCT-90

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Sallwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: MASO.C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904)375-8100

TELEFAX: (904)372-5800

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3522 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Bacillus thuringiensis

STRAIN: aizawai

INDIVIDUAL ISOLATE: PS81A2

IMMEDIATE SOURCE:

LIBRARY: Lambdaem - 11 (tm) Library of August Sick

CLONE: 81A2

US-08-291-368-1

alignment_scores:

Quality: 81.00 Length: 347

Ratio: 0.529 Gaps: 19

Percent Similarity: 44.092 Percent Identity: 19.885

alignment_block:

US-09-462-846-2 x US-08-291-368-1 ..

Align seg 1/1 to: US-08-291-368-1 from: 1 to: 3522

54 MetTyrLysGlyPheThrLeuSerGluLeuTrpClnHisHisArgHisLe 70

2173 TTGTATCAAAAATGATGATGATCAAAATTAAACCTATAACCGTTAC.. 2220

70 upheGlyInleuGluGlyAspArgPheProLeuLeuThrLysIleLeuA 87

2221CAATTAGAGG.....TATATCG 2239

87 spAlaAspGlnAspLeuSer.....ValGlnValHisProAsnAsp 100

2240 AAGATAGTCAAGACTTAGAAATCTATTATTCGTCACATCAAGCAAAACAC 2289

101 GluTyrAlaAsnIleHisGluAsnGlyGluLeuGlyLysThrGluCysTr 117

2290 GAACACAGTAATATGATACAGGATGAGGTTCCCTA.....TG 2324

117 P TyrIle..... 119

2325 CCGCGTTTCAGTCGAAGTCCATTCGATGAGAGTGTGAGAACCAATCGGT 2374

120IleAspCysGln...LysAsp 125

2375 GTGGCCACACCTTGAAATGGAATCCGTATTAATGATGTTGCTCCAGAGAC 2424

126 AlaGluIleIleTyrGlyHisAsn..... 133

2425 GGGGAAAAATGTGCACATCATTCCTCCATTCCTCGACATTCATGT 2474

134AlaThrThrLysGluGluLeuThrThrMetIleGluArgGlyLys 149

2475 TGATGACACAGACTTGCAAGAG.....GATCTAGGCGGTG 2509

149 rPAspGluLeuLeuArgValLysValLysProGluAspPhePheTyr 165

2510 GG.....GTTGTATTCACATTAAAGCAGCAAGAGGT.....TAT 2544

166 ValProSerGlyThrValHis.....AlaIleGlyLys 176

2545 GCAGATTAGCAAAATCGAATTTATTCAGAGAAACCATTAATTGAGAG 2594

176 sGlyLys..... 178

2555 AGCACTGTCTCGTGTGAGAGACGGGAAAAAATGAGAGACAAACGGG 2644

179LeuAlaLeuGluThr.....GlnGlnAsn 186

2645 AAAAATCATTCATGGAACAAACAGCATATATACAGAGCAAAAGAGCT 2694

187 SerAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspIleGlu 203

2695 GTGATGCTTATTCTGATGATTCATCAATATGATAGATTACAAACCAATAC 2744

203 LysLeu.....ArgGluLeuHisLysLys 213

2745 AAACATGCTATGATCATCGCCGACAGATAGACTTGCATCAGATCCACG 2794

213 erIleGluValIleGluValProSerIleProGluArgHisThrValHis 229

2795 AGCTTATCTTCCAGAACTACTTTCATTCACAGAAATAAATGATGGATTT 2844

230 HisGluGlnIleGluAspLeuLeuThrThrLeu..... 241

2845 TTGAGAGATATAGAAAACCGTATTTCTAGTCATTAATCCCTATATGATGC 2894

242IleGluCysAlaTyrPheSerValGlyLys.....TyrA 253

2895 GAGAAATGTCATTAATAAATGGCAATTCATTAATGCGTTATATCGCTGCGA 2944

253 snLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeu 269

2945 ACGTGAAGGCGATGATGATGATAGTACAAACAAACCAACCGCTGCGTC 2994

270 SerValIle...GluGlyGluGlyArgMetIleSerGlyLysTyrVal 285

2995 CTGTGTCCCGGATGGAGAGAGTGTCAACAAATTCGTGTGTG 3044

285 rProPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGluP 302

3045 TCCGGGGCGTGTATATCTCCGTGTATACAGCGTAC..... 3081


```

||||: ||| |||:||||: |||
2995 CTGTTGTCCCGAATGCGAAGCAGAACTGTCACAAACAAATCGTGTCTG 3044
285 rProphelyslysglyAspHisMetLeuProTyrGlyLeuGlyGluP 302
||| :|||
3045 TCCGGGGCGGTGCTATCTCCGTGTACAGCGTAC..... 3081
302 helysLeuGluGlyTyrAlaGlu...CysIleValSerHis 314
|||||:|||||
3082 .....AAAGAGGAGATGAGAGAGTTGCGTAAACCATCCAT 3117

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 14:47:19 ; Search time 128.3 Seconds
(without alignments)
4299.870 Million cell updates/sec

Title: US-09-462-846-1

Perfect score: 945
Sequence: 1 atgcagactgacgcgttatt.....aatctatcgctccatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SID56/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	945	100.0	945	20	X25733	B. subtilis cystein
2	319	33.8	945	20	X25734	B. subtilis cystein
3	311.8	33.0	948	20	X25735	B. subtilis cystein
4	258.4	27.3	12438	20	X13110	Enterococcus faeca
5	226	23.9	25002	19	V52181	Streptococcus pneu
6	225.8	23.9	2111	18	T80389	Staphylococcus aur
7	221	23.4	9623	18	V74477	Staphylococcus aur
8	175.2	18.5	3759	18	V74326	Staphylococcus pneu
9	37.2	3.9	13425	19	V52284	Streptococcus pneu
10	34.4	3.6	1082138	21	F22305	Arabidopsis thalia
11	33.8	3.6	6582	18	T99451	S. carnosus nitrat

12	33.2	3.5	3475	12	Q15172	Human erythrocyte
13	33	3.5	4040	20	Z34275	Human PKO1012 nucl
14	33	3.5	4040	21	C78573	Human PKO1012 (UNG)
15	33	3.5	7458	21	A70106	Plasmodium falcipa
16	33	3.5	1038602	20	Z01425	Complete genome se
17	32.8	3.5	1956	21	A27325	Impatiens balsamia
18	32.6	3.4	249	21	C18692	Human secreted pro
19	32.6	3.4	1448	18	V75020	Staphylococcus aur
20	32.4	3.4	1380	21	Z34650	Human growth facto
21	32.2	3.4	495	19	V43576	Human blue-light p
22	32	3.4	1507	18	V74672	Staphylococcus aur
23	32	3.4	116277	20	X20249	Borrelia burgdorfe
24	32	3.4	101715	20	X20248	Borrelia burgdorfe
25	32	3.4	1664976	19	V21209	Methanococcus jann
26	31.8	3.4	1727	15	O86279	Nitriase enzyme c
27	31.6	3.3	1150	21	C33635	Arabidopsis thalia
28	31.6	3.3	6340	18	T48871	Iams expression ve
29	31.6	3.3	7208	18	T48870	Iamh expression ve
30	31.4	3.3	1701	21	A87757	Human secreted pro
31	31.4	3.3	1816	21	A75761	cDNA encoding a hu
32	31.4	3.3	1816	21	Z52035	cDNA encoding inte
33	31.4	3.3	1816	21	Z35746	Human interleukin
34	31.4	3.3	1918	21	A75772	cDNA encoding a hu
35	31.4	3.3	1918	21	Z52046	cDNA encoding inte
36	31.4	3.3	2080	21	A95789	Human immune syste
37	31.2	3.3	1521	20	X13135	Enterococcus faeca
38	31	3.3	106746	21	A10225	Human PCTA-1 genom
39	30.8	3.3	847	20	Z16398	Human gene express
40	30.8	3.3	1118	21	C37612	Arabidopsis thalia
41	30.8	3.3	2890	19	X14100	H. pylori GHP0 282
42	30.6	3.2	663	19	C16899	Human secreted pro
43	30.6	3.2	663	19	V52432	Streptococcus pneu
44	30.6	3.2	1014	20	X13121	Enterococcus faeca
45	30.6	3.2	1742	21	F15398	Human prostate can

ALIGNMENTS

RESULT 1
ID X25733 standard; DNA: 945 BP.
AC X25733:
XX
XX 21-MAY-1999 (first entry)
DT
XX
XX B. subtilis cysteine protease Cpl coding sequence.
DE
XX
XX Cysteine protease; Gram-positive bacteria; mutation: deletion;
KW proteolysis; expression; heterologous protein; ss.
XX
OS Bacillus subtilis.
XX
XX WO9904016-A2.
PN
XX 28-JAN-1999.
PD
XX
XX 14-JUL-1998; 98WO-US14529.
PE
XX 15-JUL-1997; 97EP-0305227.
PR
XX
XX (GENV) GENENOR INT BV.
PA (GENV) GENENOR INT INC.
XX
XX Estel1 DA:
XX WPI: 1999-132260/11.
XX P-PSDB: W99371.
DR
XX Gram-positive bacterium with mutated or deleted gene for cysteine
PT protease 1, 2 or 3 - used to express proteins with reduced
PT proteolytic degradation, e.g. proteins; growth factors or enzymes

XX Disclosure; Fig 1A-B: 31bp; English.

PS This sequence represents the Bacillus subtilis cysteine protease CP1
 CC coding sequence. The invention relates to new gram-positive bacteria,
 CC especially Bacillus, containing mutations in, or deletion of, at least
 CC part of the gene encoding cysteine protease (CP1, CP2 (X25734) or CP3
 CC (X25735), so that CP-related proteolytic activity is lost. The bacteria
 CC are used for expressing heterologous proteins, particularly hormones,
 CC growth factors, cytokines or especially enzymes, e.g. protease,
 CC carboxypeptidase, lipase, isomerases (racemase, epimerase, tautomerase or
 CC mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
 CC soaps, dishwashing compositions, contact lens cleaners or laundry
 CC detergents, also for peptide hydrolysis, waste or textile treatment,
 CC to cleave fusion proteins and as animal feed additive.

XX Sequence 945 BP; 288 A; 180 C; 232 G; 245 T; 0 other:

Query Match 100.0%; Score 945; DB 20; Length 945;
 Best Local Similarity 100.0%; Pred. No. 1e-296;
 Matches 945; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 atgacgactgaacgcgttatatttcaagcctgtttcaagaagaagaatttggcgagacc 60
Db 1 atgacgactgaacgcgttatatttcaagcctgtttcaagaagaagaatttggcgagacc 60
Oy 61 gcttaagctgatttggctataccatccgttcaacagcagggagtgctggcttt 120
Db 61 gcttaagctgatttggctataccatccgttcaacagcagggagtgctggcttt 120
Oy 121 gccgcgacataaattggtcaagcgtgtgttcaaaacgggaatgataagggttcagctc 180
Db 121 gccgcgacataaattggtcaagcgtgtgttcaaaacgggaatgataagggttcagctc 180
Oy 181 agcgaattatgggaacatccacagcattatccgagcagcgttgaaggagacgcttccct 240
Db 181 agcgaattatgggaacatccacagcattatccgagcagcgttgaaggagacgcttccct 240
Oy 241 ctgcttacaanaaatatagaatgtcgcagcagactatctgttcaggttcacacgaatgat 300
Db 241 ctgcttacaanaaatatagaatgtcgcagcagactatctgttcaggttcacacgaatgat 300
Oy 301 gaatatgcacacatacattgaanaacggttgagcttggaaaaagaaatgctgttaccattat 360
Db 301 gaatatgcacacatacattgaanaacggttgagcttggaaaaagaaatgctgttaccattat 360
Oy 361 gattgcacaaaagaatgcgcagatatttatggcccaaatgacacacaaagaagaacta 420
Db 361 gattgcacaaaagaatgcgcagatatttatggcccaaatgacacacaaagaagaacta 420
Oy 421 actacacatgaagaagcgttgagaatggaatgagctcttgcgcggtgataaagtaaacgcg 480
Db 421 actacacatgaagaagcgttgagaatggaatgagctcttgcgcggtgataaagtaaacgcg 480
Oy 481 ggggaattttctatgtgcacaaagcgttactgttcaatgcatggaagaagaattcttgcct 540
Db 481 ggggaattttctatgtgcacaaagcgttactgttcaatgcatggaagaagaattcttgcct 540
Oy 541 ttggaagcgcagcagaactcagacacaaactacagattatgatattagaccgaaaagat 600
Db 541 ttggaagcgcagcagaactcagacacaaactacagattatgatattagaccgaaaagat 600
Oy 601 gcaagaaggcagcgtgcgcgagcttctatctgaaaaaagcagcattggaatgagatgcgcg 660
Db 601 gcaagaaggcagcgtgcgcgagcttctatctgaaaaaagcagcattggaatgagatgcgcg 660
Oy 661 tctattccagaagcgcatacagttcaccaatgaaatgagatttcttacaacgaca 720
Db 661 tctattccagaagcgcatacagttcaccaatgaaatgagatttcttacaacgaca 720
Oy 721 ttgattgaatgcgttacttcttcggtggggaatggaactatcaggttcagcaagctta 780
Db 721 ttgattgaatgcgttacttcttcggtggggaatggaactatcaggttcagcaagctta 780

```

```

Db 721 ttgattgaatgcgttacttcttcggtggggaatggaactatcaggttcagcaagctta 780
Oy 781 aagcagcaaaaacacatctcttctatcagttgattgaaggagggcgctatgactct 840
Db 781 aagcagcaaaaacacatctcttctatcagttgattgaaggagggcgctatgactct 840
Oy 841 ggtgaatgtctatctcttcaaaaagagatcatatgttgcgtccttaccgcttga 900
Db 841 ggtgaatgtctatctcttcaaaaagagatcatatgttgcgtccttaccgcttga 900
Oy 901 gaatttaactcgaagatatacgaatgatatgtcttcccatctg 945
Db 901 gaatttaactcgaagatatacgaatgatatgtcttcccatctg 945

```

RESULT 2
 ID X25734
 XX X25734 standard; DNA; 945 BP.

AC X25734;

DT 21-MAY-1999 (first entry)

DE B. subtilis cysteine protease CP2 coding sequence.

KW Cysteine protease; Gram-positive bacteria; mutation; deletion;
 KW proteolysis; expression; heterologous protein; ss.

OS Bacillus subtilis.

PN WC09904016-A2.

PD 28-JAN-1999.

PP 14-JUL-1998; 98WO-US14529.

PR 15-JUL-1997; 97EP-0305227.

PA (GEMV) GENENCOR INT BV.
 PA (GEMV) GENENCOR INT INC.

PI Estell DA;

DR WP1: 1999-132260/11.

DR P-PSDB; M99372.

PT Gram-positive bacterium with mutated or deleted gene for cysteine
 PT protease 1, 2 or 3 - used to express proteins with reduced
 PT proteolytic degradation, e.g. proteins, growth factors or enzymes

PS Disclosure; Fig 5A-B: 31bp; English.

CC This sequence represents the Bacillus subtilis cysteine protease CP2
 CC coding sequence. The invention relates to new gram-positive bacteria,
 CC especially Bacillus, containing mutations in, or deletion of, at least
 CC part of the gene encoding cysteine protease (CP1 (X25733), CP2 or CP3
 CC (X25735), so that CP-related proteolytic activity is lost. The bacteria
 CC are used for expressing heterologous proteins, particularly hormones,
 CC growth factors, cytokines or especially enzymes, e.g. protease,
 CC carboxypeptidase, lipase, isomerases (racemase, epimerase, tautomerase or
 CC mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
 CC soaps, dishwashing compositions, contact lens cleaners or laundry
 CC detergents, also for peptide hydrolysis, waste or textile treatment,
 CC to cleave fusion proteins and as animal feed additive.

XX Sequence 945 BP; 290 A; 189 C; 223 G; 243 T; 0 other:

Query Match 33.8%; Score 319; DB 20; Length 945;
 Best Local Similarity 59.4%; Pred. No. 2.2e-93;
 Matches 560; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

Oy 4 acgactgaacgcgttatatttcaagcctgtttcaagaagaagaatttggcgagaccgc 63

```

Db      1 atgacgcatacatatttttagagcctgtctttaaagaagaacatctggggagcgaaag 60
Qy      64 ttagctga---tttggtataccatccgcgcacagaagggagctgtgcttt 120
Db      61 ctctgtagcgtcttggcagcaataccctcaaaaaaacggtgagctgtggcgct 120
Qy      121 gcgcgcatacaaatgycacaaagcgttgtcaaaacggaatgataaagggttcaagctc 180
Db      121 tctgcaatgccatgctgcgtcgtctgtaaaaaatgcccgcgtgcaggaagacactt 180
Qy      181 aagcgaattatgggaacatccacagacattatctgcagacgcttgaaggagaccttccct 240
Db      181 gatcaagtaatggaaatccatccagagatattcgggttccggatgacaaagtgcttccg 240
Qy      241 cctgctacaanaataattagatgctgacagacattatctgtcaggtgacatccgaatgat 300
Db      241 ctgctggtataaagctgctgacgcgaatatgatactctctcgtcgaagtcacatctgtat 300
Qy      301 gaatatgccacatacatatgaanaaagctgtgagcttggaaaaaacagaatgctgtacattat 360
Db      301 gattatgcaaaaactgcacgaaatgycgacctgtgtaaacggaagtgctgtatcatct 360
Qy      361 gattgccaanaaagatgcgagattttatggcacaatgacgaacaaagaagaacta 420
Db      361 gatgcaaaagatgacgcgcaactaatcttggagacatcatgcaagcaaaagaagatgc 420
Qy      421 actacatgatagacgctgtgagaatggaatgagctcttgcgcctgttaaaagtgagcg 480
Db      421 aaacacagatagaagaagcggtgattggaacggtgctgtcgtgagcgaaatcaaaatcagca 480
Qy      481 ggggattttttctatgtgtccaaagcggttactgtctatgcatggaattggaagaattctgt 540
Db      481 gggagattcttcttattgtgcgaagcggttacacccatgcttattgtaaggaaacctgtcc 540
Qy      541 ttggaagcgaagcagaactcaagacacactcaagatlatatgatgatgacggaagaat 600
Db      541 ctggaatccagaacaaactctgtacacatattcgtatcagattatgacgcgtgtaat 600
Qy      601 gcaagaagcgaagctgcgagactctcatctgaaanaagatggaatgtagagtcggc 660
Db      601 gaccagaagccaaaaaagaactcttcatatagaanaagcattgaaagtcatacagatacga 660
Qy      661 tctatccagaacgcatagatgtaacatgtaacgaattgagattgtttacaacgaca 720
Db      661 catatcgataaagtgcataacccggaagtaaaagaattgtagacgtgagatcatgtt 720
Qy      721 ttgattgaatgcgttactcttctcgtgtgggaaatggaacttacaagatcagaagctta 780
Db      721 tatgtgaatcagatattatctcagtgtaacaatggaagattagcgcgagctgtctt 780
Qy      781 aagcagaanaaacatctctcttcatcagtgatgtgaaggaggagggccgtatgactct 840
Db      781 ccttcatataaaccctattctgtgggagtgcttctgagcggatcagaagcaaatcaaat 840
Qy      841 ggtgagatgcttcatcttcaaaaagagatcatatgtgtgctgcttcaagcttctgga 900
Db      841 aatggttcaagatgtaatgcaatgcagctcacaacttcttccctgctgagcttttga 900
Qy      901 gaattaaactcgaaagatatgcaagaatgatacgtctccacatc 943
Db      901 gaattacaatagaagaacatgtaattcatgatcatctc 943

```

RESULT 3
 X25735 X25735 standard; DNA; 948 BP.
 AC X25735;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE B.subtilis cysteine protease CP3 coding sequence.

```

XX      Cysteine protease; Gram-positive bacteria; mutation; deletion;
KW      proteolysis; expression; heterologous protein; ss.
KW      Bacillus subtilis.
OS      Bacillus subtilis.
PN      W09904016-A2.
PD      28-JAN-1999.
XX      14-JUL-1998; 98MO-0514529.
PF      15-JUL-1997; 97EP-0305227.
PR      (GENV) GENENCOR INT BV.
PA      (GENV) GENENCOR INT INC.
XX      Estrel DA;
PI      WPI: 1999-132260/11.
DR      P-PSDB; W99373.
XX      Gram-positive bacterium with mutated or deleted gene for cysteine
PT      protease 1, 2 or 3 - used to express proteins with reduced
PT      proteolytic degradation, e.g. proteins, growth factors or enzymes
XX      Disclousure; Fig 6A-B; 31pp; English.
PS      This sequence represents the Bacillus subtilis cysteine protease CP3
XX      coding sequence. The invention relates to new Gram-positive bacteria,
CC      especially Bacillus, containing mutations in, or deletion of, at least
CC      part of the gene encoding cysteine protease (CP)1 (X25733), CP2 (X25734)
CC      or CP3, so that CP-related proteolytic activity is lost. The bacteria
CC      are used for expressing heterologous proteins, particularly hormones,
CC      growth factors, cytokines or especially enzymes, e.g. trypsinase,
CC      carboxylase, lipase, isomerases (racemase, epimerase, tautomerase or
CC      mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
CC      soaps, dishwashing compositions, contact lens cleaners or laundry
CC      detergents, also for peptide hydrolysis, waste or textile treatment,
CC      to cleave fusion proteins and as animal feed additive.
XX      Sequence 948 BP; 286 A; 192 C; 244 G; 226 T; 0 other;
SQ

```

Query Match 33.0%; Score 311.8; DB 20; Length 948;
 Best Local Similarity 58.8%; Pred. No. 47e-91;
 Matches 557; Conservative 0; Mismatches 387; Indels 3; Gaps 1;

```

Qy      1 atgacgactgaacgcttattttcaagcctgtttcaagaagaattggggcggaacc 60
Db      1 atgacgcaatccacgatttttctaagcctgttgttaaagaanaaaatctggggcggaacc 60
Qy      61 gctttagctgat---ttggctataccatccgttcaacgaacgaacgaggaatgctggct 117
Db      61 gctttagagatagatttggatacagatcttcttaagatcaacgaagggagatgctggc 120
Qy      118 ttggcgcgcatcaaatatgltcaaaagcgttgttcaaaacggaatgataaagggttcaag 177
Db      121 atttcgcctacaaagaagcagcagctgtgtaaatgagccgtataaagaagaaga 180
Qy      178 ctaagcgaattatgggaacatacagacattatctgcagacgcttgaaggggagcgcctt 237
Db      181 ttgatcgaacttgggaagaagacgctggaagtatctgcggcgtagaaggagatcggctt 240
Qy      238 cctgtctacaanaataattagatgctgacagacattatctgtcaggtgacatccgaat 297
Db      241 cggctcttcaaaaagctgtgtgtagtgaagaagatacgtcaataaagttcacctgat 300
Qy      298 gatgaatatgccacaatcatatgaanaaagcttgagcttggaanaaacgaatgctgtatcatt 357
Db      301 gattactatgcggaagaagaagaggaactcgcgaagcgaaggaatgctgtgatact 360
Qy      358 attgatgtccaaaagaatgctcgagattattatgtgccacaatgcaacaacaaagaaga 417

```


Query Match	23.9%	Score 225.8	DB 18	Length 2111
Best Local Similarity	62.3%	Pred. No. 6.3e-63		
Matches 404	Conservative 0	Mismatches 238	Indels 6	Gaps 3
QY	7	actgaacggtatatttttcaagccgtgttttcaagaagaatttgggcgagaccgtta	66	
Db	676	ACTATGGCATTTATTTTATCAACCCGTTTAAAGACAGAAATTTGGGGGCTCATGCATTA	617	
QY	67	gcgcatttggcgtataccattccgtcacacgaacagagggaagtgcgtgtttgcgcg	126	
Db	616	AAAGCATTTAACTATGATGATATCCCTAATGAACCAACTGGTGAATCTGGCAATTTCTGCA	557	
QY	127	catcaaatggttgcacaaagcgtgttttaaaacggaaatgatatagaaggtttcaagcttcacgcga	186	
Db	556	CATCCAAATGGTCTTAATTAACAATTAATTCATGCTCCATACAGGAATAGACCCATGATCA	497	
QY	187	ttatgggaacatcacagaacattatttcggacagcgttgaaggagaccgtttccctcgtct	246	
Db	496	CTTTGGTACACACATCGTGAAATTAATTCGATATATATTCACAGATATGTTCCCTCTGCTA	437	
QY	247	acaaaaatatatgatgcttgacacgaacttactcgttcagggtgcatccgaatttgatgat	306	
Db	436	ACTTAAAGTTTGTGATGCCCAATGATAAATTAATCTGTTCATCTCATCCAGATGATGACTAC	377	
QY	307	gccacaatcatcatgaanaagcgtgagcgttggaaaaaacgaatgcctcgtgtacattatgtatgc	366	

Accession	Gene	Strain	Location/Qualifiers
D8	376	GCCTTTAAAAACAGAA--GCGCACTAGCTAAACAGAAATGTTGGTATAT--TTAGATGC	322
Qy	367	caaaagaatgccgagattat-ttatggccaatatgcacaaacaaaggaataactaac	425
D8	321	ACCCACAGGTGCGAATTAATTAATTATGTCGCTCATGCACATTAATTAACAGCGTTAAATGCA	262
Qy	426	catgatagacgttggagaatgagatgagctcttcgcgcgttaagtaagccggagggga	485
D8	261	CATGATGATATATCATGTAATTTGGCAGCTTTTCAACGATATACGTTAAGCCCGGTGA	202
Qy	486	tttttcctatgtcgaacgagctactgtcttcatacgcatgttgaaagaatcctgtcttga	545
D8	201	CTTTATTAATGATGCTCTGCTGGCACTGTCCATGCGCATTTGGTTCGGGATATTTTAATTTTACA	142
Qy	546	gacgcagcagaactaagaacacaaactacagatataatgatatgaccgaagaatgcaga	605
D8	141	GACACAAATCATTCGCTACGATACCATTCGATTTAAGTATGATAGACGATCATANA	82
Qy	606	aggcaagctgcgcgagcttcattcctcgaaaaagacattgaagtatga	653
D8	81	TGGTCAATTTGGCTGATTTGGCACTTAGCAAGAAAGTAAGATGATTTGA	34
RESULT 7			
ID	V74477/c		
AC	V74477	standard; DNA; 9623 BP.	
XX			
DE	16-MAR-1999	(first entry)	
XX			
DE	Staphylococcus aureus contig SEQ ID #166.		
XX			
XX	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KW	skin infection; surgical wound infection; scalded skin syndrome;		
KW	toxic shock syndrome; ds.		
XX			
OS	Staphylococcus aureus.		
XX			
XX	Key		
FT	misc-feature	Location/Qualifiers	
FT		361..420	
FT		/*tag= a	
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc-feature	2161..2220	
FT		/*tag= b	
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc-feature	3961..4020	
FT		/*tag= c	
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc-feature	5761..5820	
FT		/*tag= d	
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc-feature	7561..7620	
FT		/*tag= e	
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc-feature	9361..9420	

/*tag= f
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP786519-A2.
30-JUL-1997.
07-JAN-1997; 97EP-0100117.
05-JAN-1996; 96US-0009861.
(HUMA-) HUMAN GENOME SCI INC.
Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
Rosen CA;
WPI: 1997-374922/35.
Polynucleotide(s) and proteins derived from *Staphylococcus aureus* stored on computer readable medium and used in the production of anti-*S. aureus* vaccines
Claim 1; Page 816-822; 3271pp; English.
This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the *S. aureus* DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against *S. aureus* infection. The polypeptides can also be used in a kit for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the *S. aureus* DNA sequences contained on the computer readable medium.
Sequence 9623 BP: 3025 A; 1508 G; 1655 G; 3072 T; 363 other;

Query Match 23.4%; Score 221; DB 18; Length 9623;
Best Local Similarity 50.6%; Pred. NO. 5.3e-61;
Matches 474; Conservative 0; Mismatches 460; Indels 3; Gaps 1;

QY 7 actgaaccgttatttttcaagcctgttttcaagaagaatttggcgcggaaccgttla 66
DB 6142 ACTATGCGATATTTTACACCGCTTTTAAAGACAGAAATTTGGCGGTCATGCACTTA 6083
QY 67 gctgttttgctatcaccttcgcgaacgaagcgaggtgtgtgtgtgtgtgtgtgt 126
DB 6082 AAAGCATTTAACTATGATATCCCTTAATGAACACGTGTGATGCTGGCAATTTTCGA 6023
QY 127 catcaaatggtcgaagcgtgtgttcaaaacgaatgtatataaggggttcacgctcaagaa 186
DB 6022 CATCCAAATGCTCTTAATACATATATCAATGTCATACAGATGATGACCTTGATCA 5963
QY 187 ttaaggaaactcacagacatttttggagacgttgaaggcgcttccctcgtgtc 246
DB 5962 CTTTGCTCAACATCGGATTTGATTAATGATTCACAGATGATGTTCCCTGCTGA 5903
QY 247 acaaaaattatagatcgtcgaagactatcgttcaggtgcacccgaatgatgaat 306
DB 5902 ACTAAAGTATATGATGCCAATGATTAATTAATCTGTCAAGTTCAACCAAGATGATGACTAC 5843

QY 307 gccacatcacatgaaaaacgtgagcttgcgaaaaaacgaatcgtcgtcatattgtatgc 366
DB 5842 GCTTTAAACACGAA---GGCGAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5786
QY 367 caaaaagatgcgagatatttattgcccacaatgcgaacaagaagaactaactacc 426
DB 5785 NNN 5726
QY 427 atgatatcgctgagagaatgagatgagcttcttgcgcgtgtgtaagtgaaagcgagat 486
DB 5725 ATGATTTGATTAATCATGAAATTTGACAGTCTTTTCAACGCTTACCTGTTAAGCCGCTGAC 5666
QY 487 ttctctatgccaagcgtactgtctacgtcagatttggaaaaggaaattcctgttggag 546
DB 5665 TTTTATATGTTCCGTCGTCACGTGTCATGCCATGTTGTTGGGTATTTAATTTAGAG 5606
QY 547 acgcagcgaactcgaacacaaactcacagattatgatatgaacgaagaatgtagaa 606
DB 5605 ACACACAAATCCTCGATACACATACCGTATTTATGACTATGATGACGATCAAAAT 5546
QY 607 ggcagcgtcgcgagctcatcgtcgaagaagaagcattgaagtagaagtcctgtat 666
DB 5545 GGTCAATTTGCGTGAATTTGCACTTAGACCAAAAGTAAAGTGTGATTAATTAATTAAT 5486
QY 667 ccagacgagcatcacagttcaccaatgaagaatgagattgttcaaacagacattgatt 726
DB 5485 GAACCAATTAATCTCTTATTAACACTACACACACACACACTGTAATCAATTTGTA 5426
QY 727 gaatgcgttacttcttgcgttggaagaatggaactcacagatcacgaactaaagcag 786
DB 5425 GATATTCATATTTTGGCAGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 5366
QY 787 caaaaacattcctctatcagtgatgagtgaaagcggtgtatgactcctgtgtgag 846
DB 5365 CCACATGCTTATGCTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5306
QY 847 tatgtatccttcaaaaagaagatcatatgttctgctcctaagcgttcttgaagaatt 906
DB 5305 AAGCCTATGACATCCAAAGGTTCTAGCTTATTTTAACTACGGAAGACGAGATAT 5246
QY 907 aaatcgaagatatgcagaatgtatcgtctccacc 943
DB 5245 CACTTGTGCGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5209

RESULT 8
V74326/c
ID V74326 standard; DNA; 3759 BP.
XX
AC V74326;
XX
DT 16-MAR-1999 (first entry)
XX
XX

Staphylococcus aureus contig SEQ ID #15.

Computer readable medium; vaccine; *S. aureus* infection; immunodetection; cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy; skin infection; surgical wound infection; scalded skin syndrome; toxic shock syndrome; ds.

Staphylococcus aureus.

Key Location/Qualifiers

misc_feature 421..480

misc_feature

misc_feature

misc_feature

misc_feature

2221..2280
/*tag= b
/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP786519-A2.

30-JUL-1997.

07-JAN-1997; 97EP-0100117.

05-JAN-1996; 96US-0009861.

(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

Rosen CA;

WPI; 1997-374922/35.

Polynucleotide(s) and proteins derived from *Staphylococcus aureus* stored on computer readable medium and used in the production of anti-*S. aureus* vaccines

Claim 1; Page 255-257, 3271pp; English.

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the *S. aureus* DNA sequences allows putative functions to be assigned so that the protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against *S. aureus* infection. The polypeptides can also be used in a kit for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the *S. aureus* DNA sequences contained on the computer readable medium.

Sequence 3759 BP; 1165 A; 668 C; 510 G; 1289 T; 127 other;

Query Match

Best Local Similarity

Matches

443; Conservative

1; Mismatches

489; Indels

1; Gaps

1;

Score 175.2; DB 18; Length 3759;

Pred. No. 2,4e-46;

18.5%;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

372 agatgcgagatattatgcccacatgcaacaacaagaagaactactacatgat 431

606 AGATGCAGAAATAGTTATAGGAGCATTTAGCAGAGTATAGAGAAAGTTGCCATCATGT 547

432 agagcgtgagagatgagatgagctctgcccgtglaaaglaagccggggagatltt 491

546 TCAACACGAGAGATAGAGTGCATCTGATATATTAAGTAAMCCGCGKGAATTCCTA 487

492 ctaigtccaaagcgtaactgctgcatgcatggaagaagaattccttgcttgagacgca 551

486 TTTTATNN 427

552 gcaagaactgacacaaactacagatatactatgacccgaaagaatgacagaagcaa 611

426 NNNNNNGTCAGACATTCATATACCTTATGATTTCAATCGTCAAGATTAATCAATATA 367

612 gctgcgagactcatctggaagaagaagcatggaagtgaagagccgctctatccaga 671

366 TGATGACCGTTAATATATGAAAAGCTTAGACGTTATTCAGTACATGCAATGCAATAC 307

672 acggcatagatcaccatgacaagaattgagagattgcttacaacagacattgattgaatg 731

306 TAAATATTTTGGCTGAAGCGAAATTAATGAAACATTAAGTGTACACATGTTATCGAA 247

732 cgttactcttcggtgggaatggaactatcagatcagatcagaagcttaagaagcaaaa 791

246 TGATTTCTTTCAATGTTAATGGAATTTCTGACAGTTAATTAATTAATGAAAGCTTAG 187

792 accattctctatcagtgatgagagagagcgatgacatctcgtgagatgagatg 851

186 AGAGTTCTGTTAGTACAGTGTGGAAGCGCAAAATGATGTCGAGATGGAAT 127

852 ctatcttcaaaaagagatcactatgctgcttcaagcttcttggaagaattaaact 911

126 TTTCAACTGCTCTGCTGACAACTTATTTGACTTCTGAAAGATTGATAGTGTCT 67

912 cgaagatagcagaatgctgctccatctg 945

66 TGAAGTGTATTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 33

RESULT 9

V52284/C

V52284 standard; DNA; 13425 BP.

V52284;

23-OCT-1998 (first entry)

Streptococcus pneumoniae genome fragment SEQ ID NO:151.

Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

computer readable medium; vaccine; pharmaceutical composition; ds.

Streptococcus pneumoniae.

MO9818931-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US19588.

31-OCT-1996; 96US-0029960.

(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

Kunsch CA, Rosen CA;

WPI; 1998-272225/24.

Computer-readable medium with recorded *Streptococcus pneumoniae*

PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 pneumoniae

Claim 1; Page 1009-1017; 1409pp; English.

XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
 CC to 391, identifying members of the library which contain sequences
 CC that hybridize to the target sequence and isolating the nucleic acid
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
 CC from an organism, amplifying nucleic acid molecules whose nucleotide
 CC sequence is homologous to amplification primers derived from the
 CC fragment of the S. pneumoniae genome to prime the amplification and
 CC isolating the amplified sequences. The computer readable medium can be
 CC used in a computer-based system for identifying fragments of the
 CC S. pneumoniae genome of commercial importance, or expression modulating
 CC fragments of the S. pneumoniae genome. Products from the present
 CC invention can be used in diagnosis kits and assays, and pharmaceutical
 CC compositions and vaccines for S. pneumoniae.

XX Sequence 13425 BP; 4434 A; 2362 C; 2406 G; 4223 T; 0 other;

Query Match 3.9%; Score 37.2; DB 19; Length 13425;
 Best Local Similarity 51.9%; Pred. No. 0.32; Mismatches 78; Indels 0; Gaps 0;
 Matches 84; Conservative 0;

QY 162 gtataaagggttaacgctcgaatattggaacatcacacagacattatcgagcgt 221
 Db 7527 GAATTAAGAACTCAAAATGGAAGATTATTGAAATACAAAGATATTATTAGCAATTT 7468
 QY 222 tgaaggagaccgtttccctcgtctacaacaaatattagatgctgagccaggaactatcgt 281
 Db 7467 TAAACCTAAACCTGTCATTTGCGACTACAAAGAGTAATTAATCTGCTTTTATTGG 7408
 QY 282 tcagggtcatccgaatgatgatgcacacatcatgaanaa 323
 Db 7407 TCATGATGATTTGAAATTAATTTGTCAAAGTACAGCAGCAAAA 7366

RESULT 10
 F22305
 ID F22305 standard; DNA; 1082138 BP.

XX F22305;
 AC
 XX 20-MAR-2001 (first entry)
 DT
 XX Arabidopsis thaliana chromosome 4 centromere.
 DE
 XX Centromere; microsome; vector; ds.
 KM
 XX Arabidopsis thaliana.
 OS
 PN W0200055325-A2.
 PD
 XX 21-SEP-2000.
 PD
 XX 17-MAR-2000; 2000WO-US07392.
 PF
 XX 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.
 PA
 XX Preuss D, Copenhagen G, Keith K;
 PI
 XX WPI: 2000-567529/55.
 DR

XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells

Claim 68; Page 977-1388; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;

Query Match 3.6%; Score 34.4; DB 21; Length 1082138;
 Best Local Similarity 52.0%; Pred. No. 28;
 Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 259 gatgctgaccaggaactatcgttcaggtgcacccgaatgatgatgccaacatcat 318
 Db 245148 ggtgatgagatgcattctattctcgtatcattttctctatgaaacagtcacagtcctt 245207
 QY 319 gaaacggtgagccttggaacaaacagatcgtgtagcatattggtgcacaaagatgcc 378
 Db 245208 tctaatagaaaaggaagaaaataatatttaagtaattagtcacaaaaaaaga 245267
 QY 379 gagattattatgcccacaatgcaaca 406
 Db 245268 aaatctattatgagatattatgctaaa 245295

RESULT 11
 T99451
 ID T99451 standard; DNA; 6582 BP.

XX T99451;
 AC
 XX 11-MAY-1998 (first entry)
 DT

XX S. carnosus nitrate reductase molybdenum cofactor genes.

XX Nitrite reductase molybdenum cofactor; MoEB; MoAB; MoAC; MoAB;
 KM MoAB; MoBA; MoBA; MoAB; MoAB; MoAB; MoAB; MoAB; MoAB; MoAB;
 KM pollutant; biotransformation; water treatment; vegetable; ds.
 XX Staphylococcus carnosus DSM 10563.

XX Key Location/Qualifiers
 FH -35_signal 55..60 /tag= a
 FT -10_signal 78..83 /tag= b
 FT RBS 101..115 /tag= c
 FT CDS 117..118 /tag= d
 FT /function= moeb gene
 FT 1150..1662 /tag= e
 FT /function= moab gene
 FT /function= complement (1846..2331)
 FT /tag= f
 FT /function= moac gene

```

FT CDS 2396..3655
FT /*tag= g
FT //function= moea gene
FT CDS 3652..4134
FT /*tag= h
FT //function= mobb gene
FT CDS 4131..4583
FT /*tag= l
FT //function= moae gene
FT CDS 4583..4816
FT /*tag= j
FT //function= moad gene
FT CDS 4821..5411
FT /*tag= k
FT //function= moba gene
FT CDS 5470..6492
FT /*tag= l
FT //function= moa gene
FT 6500..6543
FT terminator
FT /*tag= m

EP805205-A1.
05-NOV-1997.
02-MAY-1996; 96EP-0201224.
02-MAY-1996; 96EP-0201224.
02-MAY-1996; 96EP-0201224.
(NEST ) SOC PROD NESTLE SA.
Fast B, Galer W, Goeltz F, Lindgren P, Neubauer H;
Pantel I;
WPI: 1997-529059/49.
P-PSDB: W26741-49.

Recombinant S. carnosus nitrate and nitrite reductase proteins -
useful for reducing nitrate and nitrite in nitrate-polluted material

Claim 3; Page 42-49; 66pp; English.

This DNA sequence from staphylococcus carnosus includes genes that
code for the nitrate reductase molybdenum cofactor biosynthesis
proteins Moeb, Moab, Moac, Moaa, Moab, Moab, Moab and Moae
(see W26741-49). This gene region was identified following Tn917
insertion mutagenesis studies. Claimed DNA molecules (see
T99449-51) can be used in the recombinant production of claimed
proteins (see W26733-49) involved in nitrate and nitrite reduction
in S. carnosus. These recombinant proteins, or cells expressing
them, in free or immobilised form, can be used in a claimed method
for reducing nitrate and nitrite in nitrate-polluted water,
vegetables and other polluted materials.

Sequence 6582 BP; 2286 A; 980 C; 1343 G; 1973 T; 0 other;
SQ

```

Query Match 3.6%; Score 33.8; DB 18; Length 6582;
 Best Local Similarity 49.2%; Pred. No. 2.7;
 Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```

OY 233 gttccctcgtctcaaaaaatattagatgctgacagactatctgtcaggtgcatc 292
    ||||| || || || || || || || || || || || || || || || || || ||
DB 5248 gtttcgcatattatgcacgctgcttagagtcagatgactatgctgctcatgtttac 5307
    || || || || || || || || || || || || || || || || || || || ||
OY 293 cgaatgtgataatgcacacatatacagtaaaagctgagcttggaaaaacaaatgctgt 352
    || || || || || || || || || || || || || || || || || || || ||
DB 5308 aacaaacgcctctgattgatacagatgataagctagatgatacgaagaatgata 5367
    || || || || || || || || || || || || || || || || || || || ||
OY 353 acattatgattgccaagaagatgcagagattattatgcccacaaatgcacaaagaag 412
    || || || || || || || || || || || || || || || || || || || ||
DB 5368 agaataattatccccaagatttagagagacataaagaataataacacagtttaacg 5427
    || || || || || || || || || || || || || || || || || || || ||

```

OY 413 a 413
 DB 5428 a 5428

```

RESULT 12
ID Q15172 standard; cDNA; 3475 BP.
AC Q15172;
DE 08-MAR-1992 (first entry)
DE Human erythrocyte membrane anion-transport protein (band 3) gene..
DE Senescent cell peptide; ANION 1; COOH; auto immune disease.
DE Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 151..2886
FT /*tag= a

WO9118013-A.
28-NOV-1991.
24-MAY-1991; 91WO-US03557.
24-MAY-1990; 90US-0528210.
(RES ) RES CORP TECH INC.
Key MBM, Marchalonis JJ;
WPI: 1991-369184/50.
P-PSDB: R15355.

Synthetic senescent cell antigen and peptide(s) - used to detect
antibodies to SCA, study cellular ageing and auto-immune
mechanisms, separate anions and treat auto-immune disease

Disclosure: Fig 1; 74pp; English.

The sequence is that of the cDNA encoding the human erythrocyte
anion-transport protein (band 3). Nucleotides 1612-1662 (ANION 1) or
2434-2481 (COOH) of this sequence encode a peptide which is immuno-
reactive with antibodies to senescent cell antigen (SCA). The antigen
and peptides can be used in any method or technique involving or
requiring the inhibition of the binding of senescent cell IgG to
senescent cells in vitro or in vivo. The prods. can be used in to
examine various physiological, biochemical, and immunological
mechanisms in humans and other animals, such as those relating to
cellular ageing, degeneration, and life span; removal of senescent
cells; anion transport across cellular membranes; disease
mechanisms where increased senescent cell IgG may play a role, such
as haemolytic anaemia, sickle cell anaemia, and idiopathic
thrombocytopenia purpura; other autoimmune disease mechanisms, such
as rheumatoid arthritis and systemic lupus erythematosus; and other
diseases such as those involved in malaria and cancer. More
specifically, the prods. are useful in cosmetic and pharmaceutical
compos., diagnostic kits, and methods for detecting or measuring
Abs to SCA, studying cellular ageing and autoimmune mechanisms,
separating anions from a gas or liquid, or diagnosis and treatment
of certain diseases.

Sequence 3475 BP; 672 A; 1050 C; 997 G; 756 T; 0 other;
SQ

```

Query Match 3.5%; Score 33.2; DB 12; Length 3475;
 Best Local Similarity 50.6%; Pred. No. 3;
 Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI
XX WPI: 2000-365347/31.
DR
XX

XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX

PS Disclosure: Page 462-464; 577pp; English.
XX

CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX

SQ Sequence 7458 BP: 3523 A; 514 C; 1174 G; 2247 T; 0 other:

Query Match 3.5%; Score 33; DB 21; Length 7458;

Best Local Similarity 49.2%; Pred. No. 5.3;

Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 258 agatgctgaccagactatctgttcagggtgcatccgaatgataatgccacataca 317
II III III III III III III III III III III III III III III
DB 5940 agtgatgaaagaagatgatacgaagaatgagtgaggagataagaagaatgaa 5999
QY 318 tgaatacgtgtgcttggaataacagaatgctgfacattattgttgccaaagaatgc 377
II III III III III III III III III III III III III III III
DB 6000 tgggggagataagaatgaaatgaaatgagagataaaatgaggaataatggaga 6059
QY 378 cgagattattatggccacatgcaacaagaagaagaactaactacatgataga 434
II III III III III III III III III III III III III III III
DB 6060 acataaataagaagaataaagagacataagaatgataactaataatagaagaaca 6116

Search completed: May 17, 2001, 15:41:11
Job time: 3232 sec

•
•
•
•
•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 14:52:15 ; Search time 67.19 Seconds
(without alignments)
2455.716 Million cell updates/sec

Title: US-09-462-846-1

Perfect score: 945
Sequence: 1 atgaagcactgaaccgtatt.....aatgtatcgtctcccatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 segs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCRNUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	50	5.3	7218	1	US-08-232-463-14 Sequence 14, Appl
C 2	32.2	3.4	495	4	US-08-964-268-19 Sequence 19, Appl
C 3	30.6	3.2	3600	3	US-08-894-731-1 Sequence 1, Appl
C 4	30.2	3.2	10299	2	US-08-477-451-1 Sequence 1, Appl
C 5	30.2	3.2	10299	2	US-08-477-451-5 Sequence 5, Appl
C 6	30.2	3.2	19932	2	US-08-477-451-25 Sequence 25, Appl
C 7	29.8	3.2	3128	4	US-08-716-449-1 Sequence 1, Appl
C 8	29.6	3.1	2185	4	US-08-714-918-83 Sequence 83, Appl
C 9	29.6	3.1	2185	4	US-09-265-315-83 Sequence 83, Appl
C 10	29.6	3.1	2185	4	US-09-265-315-83 Sequence 83, Appl
C 11	29.2	3.1	1549	2	US-08-317-880-5 Sequence 5, Appl
C 12	29.2	3.1	1549	2	US-08-783-396-5 Sequence 5, Appl
C 13	29.2	3.1	2471	1	US-08-152-271-1 Sequence 1, Appl
C 14	29.2	3.1	3954	1	US-08-317-880-1 Sequence 1, Appl
C 15	29.2	3.1	3954	1	US-08-782-396-1 Sequence 1, Appl
C 16	28.8	3.0	1448	1	US-08-585-758A-1 Sequence 1, Appl
C 17	28.8	3.0	1448	1	US-08-977-818-1 Sequence 1, Appl
C 18	28.8	3.0	1448	2	US-08-670-274B-1 Sequence 1, Appl
C 19	28.8	3.0	12720	2	US-08-403-866-11 Sequence 11, Appl
C 20	28.8	3.0	16442	3	US-08-781-891-208 Sequence 208, App
C 21	28.6	3.0	1829	6	US-08-687-080-57 Sequence 57, Appl
C 22	28.6	3.0	2214	6	US-08-5258502-1 Patent No. 5258502
C 23	28.4	3.0	2256	2	US-08-896-590A-3 Sequence 3, Appl
C 24	28.4	3.0	2256	2	US-08-896-590A-1 Sequence 1, Appl
C 25	28.4	3.0	12047	2	US-09-022-461-1 Sequence 1, Appl
C 26	28.2	3.0	1244	4	US-08-591-685-12 Sequence 12, Appl
C 27	28.2	3.0	2479	4	US-09-050-863-4 Sequence 4, Appl

C 28	28.2	3.0	5022	2	US-09-289-368-1 Sequence 1, Appl
C 29	28.2	3.0	5183	3	US-09-039-555B-18 Sequence 18, Appl
C 30	28.2	3.0	5243	2	US-08-414-335-2 Sequence 2, Appl
C 31	28	3.0	1209	4	US-08-822-774-18 Sequence 1, Appl
C 32	28	3.0	7400	4	US-07-674-852-1 Sequence 1, Appl
C 33	28	3.0	7400	4	US-08-473-185-1 Sequence 5, Appl
C 34	28	3.0	9171	1	US-08-038-682-5 Sequence 5, Appl
C 35	28	3.0	9171	1	US-08-302-832-5 Sequence 5, Appl
C 36	28	3.0	9171	2	US-08-530-198-5 Sequence 5, Appl
C 37	28	3.0	9171	2	US-08-469-880-5 Sequence 5, Appl
C 38	28	3.0	9171	2	US-08-728-470-5 Sequence 5, Appl
C 39	28	3.0	9171	2	US-08-617-697-5 Sequence 5, Appl
C 40	28	3.0	9323	1	US-08-038-682-6 Sequence 6, Appl
C 41	28	3.0	9323	1	US-08-302-832-6 Sequence 6, Appl
C 42	28	3.0	9323	2	US-08-530-198-6 Sequence 6, Appl
C 43	28	3.0	9323	2	US-08-469-880-6 Sequence 6, Appl
C 44	28	3.0	9323	2	US-08-728-470-6 Sequence 6, Appl
C 45	28	3.0	9323	2	US-08-617-697-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)863-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

```

Query Match      5.3%; Score 50; DB 1; Length 7218;
Best Local Similarity 3.0%; Pred. No. 5e-06;
Matches 11; Conservative 208; Mismatches 143; Indels 0; Gaps 0;

QY 294 gaatgtagatgccaacatcatgaaacggttgagcttgaaaaacagaatgctgtgta 353
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 1452 GATAGAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1393

QY 354 catlatactgccaacagatccgagattattatggccaacagacacacagaaga 413
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1392 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1333

QY 414 agaacactaactacatgatagagcgtggaatggaatgagctctgcgcgtgtgtaagt 473
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1332 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1273

QY 474 aagccgggggagtttttctatgctgccaagcggtgctgtctatgagatggaaagat 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1272 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213

QY 534 tcttgcttgagagcgcagacgaactcacacacactacatattatgattatgacgcg 593
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1212 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1153

QY 594 aaaagatgcagaagcgaagctgcgcgagctcatctgaaaaagagcatgaaagtataga 653
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1152 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1093

QY 654 gg 655
    : :
Db 1092 RR 1091

RESULT 2
US-08-964-268-19/c
; Sequence 19, Application US/08964268
; Patent No. 6114503
; GENERAL INFORMATION:
; APPLICANT: WEI, YING-BEI
; APPLICANT: RUBEN, STEVEN M
; APPLICANT: SANCAR, AZIZ
; APPLICANT: HSU, SHIAO-WEN D
; APPLICANT: KAZANTSEV, ALEXSEY G
; TITLE OF INVENTION: HUMAN BLUE-LIGHT PHOTORECEPTOR hCRY2
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,268
; FILING DATE: HERewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,189
; FILING DATE: 04-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0750001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-964-268-19

Query Match      3.4%; Score 32.2; DB 4; Length 495;
Best Local Similarity 47.3%; Pred. No. 0.64;
Matches 61; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 318 tgaacacggtgagcttggaacacagaatgctgttacattatgattgccaagaatgctg 377
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 426 TGAANAAGTTGAACNGGMAANGNNTGCCATGCCCTAAGNNTNGGAAAAATTANNC 367

QY 378 cgagattattatgcccacatgcaacacacagaagaagaaactaactacatgataagcg 437
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 366 CGAAGTTGTCNTGCCAACNTNTAACCTNAGGGAAGAAAGAACNCCANACAGNC 307

QY 438 tggagaatg 446
    || || ||
Db 306 AGGAAAAG 298

RESULT 3
US-08-894-731-1
; Sequence 1, Application US/08894731
; Patent No. 6084089
; GENERAL INFORMATION:
; APPLICANT: MINE, TOSHIKI
; APPLICANT: OHYAMA, AKIO
; APPLICANT: HIYOSHI, TORU
; APPLICANT: KASAKURA, KEISUKE
; TITLE OF INVENTION: COLD-INDUCIBLE PROMOTER SEQUENCE
; FILE REFERENCE: 760-234P
; CURRENT APPLICATION NUMBER: US/08/894,731
; CURRENT FILING DATE: 1997-10-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Solanum tuberosum
US-08-894-731-1

Query Match      3.2%; Score 30.6; DB 3; Length 3600;
Best Local Similarity 49.7%; Pred. No. 6.5;
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 216 acagcttgaaggagacgttccctctgcttacaacaaatattagatgctgaccagactt 275
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 156 aaaaattgtgtactattatttatttatttatttatttatttatttatttatttattt 215

QY 276 atctgttcaggtgcatccgaatgataatatacgcaacatacatgaaacggtgaactgg 335
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 216 aaatgggttgacatcactacataaataatgaaaaatcgacgataagtgagctctaa 275

QY 336 aaaaacagaatgctgttacattatgattgccaanaa 372
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 276 aaattgacaatgcttcattgatttttttttcatgcaaaa 312

RESULT 4
US-08-477-451-1
; Sequence 1, Application US/08477451
; Patent No. 592865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation

```


Query Match	3.28;	Score 30.2;	DB 2;	Length 10299;
Best Local Similarity	52.8%;	Pred. No. 16;		
Matches	65;	Conservative	0;	Mismatches 58; Indels 0; Gaps 0;

Db 5985 TCA 5983

RESULT 6
US-08-477-451-25

```

1  GENERAL INFORMATION:
2  APPLICANT: COVACCI, Antonello
3  TITLE OF INVENTION: Helicobacter Pylori CagI Region
4  NUMBER OF SEQUENCES: 46
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Chiron Corporation
7  STREET: 4560 Horton Street
8  CITY: Emeryville
9  STATE: CA
10 COUNTRY: USA
11 ZIP: 94608-2916
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentin Release #1.0, Version #1.30
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/477,451
21 FILING DATE: 07-JUN-1995
22
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: McClung, Barbara G.
26 REGISTRATION NUMBER: 33,113
27 REFERENCE/DOCKET NUMBER: 0335_002
28
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 510-601-2708
31 TELEFAX: 510-655-3542
32
33 INFORMATION FOR SEQ ID NO: 25:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 19932 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: single
38 TOPOLOGY: linear
39
40 MOLECULE TYPE: DNA (genomic)
41
42 US-08-477-451-25

```

```

? ORGANELLE:
? IMMEDIATE SOURCE:
? LIBRARY: lambda zap tuber cDNA
? CLONE: 1.2.1 and E2
? POSITION IN GENOME:
? CHROMOSOME/SEGMENT:
? MAP POSITION:
? UNITS:
? FEATURE: open reading frame
? NAME/KEY: starch branching enzyme
? LOCATION: 44-2788
? IDENTIFICATION METHOD: lone ORF with homology to other starch
? IDENTIFICATION METHOD: branching enzymes
? OTHER INFORMATION: complements KV832 E. coli glycogen
? OTHER INFORMATION: branching enzyme mutant
?
JS-08-716-449-1

```

```

Query Matchc 3.2%: Score 29.8; DB 4; Length 3128;
Best Local Similarity 54.0%: Pred. No. 11;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Ox 311 acatacgaataaacgcttgacgttggaaaacagaaactgctgtgataattgattgccaa 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2973 ACATATTCATGCAATGCACCTTTTCATGTACACMACCAGCAATTAATTTTGATGCAAT 2914
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2913 AATATGCAATGAGCAATATCTTAAAAAATACTATGATGCAATCACT 2661

RESULT 8
US-08-714-918-83
; Sequence 83, Application US/08714918
; Patent No. 6037123
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Filth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:

```

TELEX: 67-3510

```

; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2185 base pairs

```

Page 6

RESULT 13
US-08-152-271-1
; Sequence 1, Application US/08152271
; Patent No. 5583039
; GENERAL INFORMATION:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 11:56:34 ; Search time 1125.86 Seconds
(without alignments)
7332.717 Million cell updates/sec

Title: US-09-462-846-1

Perfect score: 945
Sequence: 1 atgacgactgaacgcgtatt.....aatgtatcgtctcccatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: em_gss_pin1:*
 191: em_gss_pin2:*
 192: em_gss_pro:*
 193: em_gss_rod1:*
 194: em_gss_rod2:*
 195: em_gss_rod3:*
 196: em_gss_rod4:*
 197: em_gss_rod5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: gb_gss1:*
 202: gb_gss2:*
 203: gb_gss3:*
 204: gb_gss4:*
 205: gb_gss5:*
 206: gb_gss6:*
 207: gb_gss7:*
 208: gb_gss8:*
 209: gb_gss9:*
 210: gb_gss10:*
 211: gb_gss11:*
 212: gb_gss12:*
 213: gb_gss13:*
 214: gb_gss14:*
 215: gb_gss15:*
 216: gb_gss16:*
 217: gb_gss17:*
 218: gb_gss18:*
 219: gb_gss19:*
 220: gb_gss20:*
 221: gb_gss21:*
 222: gb_gss22:*
 223: gb_gss23:*
 224: gb_gss24:*
 225: gb_gss25:*
 226: gb_gss26:*
 227: gb_gss27:*
 228: gb_gss28:*
 229: gb_gss29:*
 230: gb_gss30:*
 231: gb_gss31:*
 232: gb_gss32:*
 233: gb_gss33:*
 234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	48	5.1	997	229	CNS005TE	AL060767 Drosophila
C 2	41.2	4.4	1204	229	CNS016E2	AL106628 Drosophila
C 3	39.8	4.2	292	201	AO070844	AO070844 HS-2256_A
C 4	38.6	4.1	1095	232	CNS05PY4	AL348529 Tetradon
C 5	38.4	4.1	419	209	AO675159	AO675159 HS-2162_A
C 6	38.2	4.0	395	201	AA550191	AA550191 1307m3 gm
C 7	38.2	4.0	712	151	BF628576	BF628576 HVSMEB000
C 8	38.2	4.0	375	202	AO137339	AO137339 HS-3058_B
C 9	37.8	4.0	635	169	BF815061	BF815061 MR2-C1012
C 10	37.6	4.0	889	217	AZ209683	AZ209683 SE-0138_A
C 11	37	3.9	799	212	AQ842470	AQ842470 CP61126A
C 12	37	3.9	1079	220	CNS02H03	AL197004 Tetradon
C 13	36.8	3.9	411	168	BF727776	BF727776 1000052F0
C 14	36.8	3.9	1101	229	CNS00GDR	AL072060 Drosophila
C 15	36.6	3.9	1101	229	CNS0039G	AL063921 Drosophila
C 16	36.4	3.9	249	26	AV231078	AV231078 AV231078
C 17	36.4	3.9	966	229	CNS00154	AL075573 Drosophila
C 18	35.8	3.8	306	127	BB200489	BB200489 BB200489

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	JOURNAL
C 19	35.8	3.8	451	15	AI071988	UI-R-C2-n			
C 20	35.8	3.8	855	229	CNS0182P	AI100556	Drosophila		
C 21	35.8	3.8	1101	229	CNS0182P	AI108811	Drosophila		
C 22	35.6	3.8	504	115	AW463593	BP250013A			
C 23	35.6	3.8	530	201	AO075795	CIT-HSP-2			
C 24	35.4	3.7	444	146	BF320729	uz55907.y			
C 25	35.4	3.7	355	20	AI408821	EST237112			
C 26	35.4	3.7	484	140	BE681867	UI-M-AQ0-			
C 27	35.4	3.7	491	214	AO992818	RPCL-23-2			
C 28	35.4	3.7	544	144	BF113427	EST440817			
C 29	35.4	3.7	576	210	AO692130	HS_5328_B			
C 30	35.4	3.7	609	167	BE460767	EST412186			
C 31	35.4	3.7	801	231	CNS04FNM	AL288581	Tetradon		
C 32	35.4	3.7	1101	229	CNS014JS	AI104242	Drosophila		
C 33	35.2	3.7	364	2	AA111946	zm53e09.T			
C 34	35.2	3.7	711	232	PT006001R	AL447600	Parameciu		
C 35	35.2	3.7	874	150	BF576179	BF576179	602132616		
C 36	35.2	3.7	1201	229	CNS0160N	AI106145	Drosophila		
C 37	35	3.7	502	208	AO586301	RPCL-11-4			
C 38	35	3.7	925	229	CNS001BN	AI074475	Drosophila		
C 39	35	3.7	1101	229	CNS012VN	AI102077	Drosophila		
C 40	35	3.7	1201	229	CNS0108G	AL098650	Drosophila		
C 41	34.8	3.7	442	164	BE222338	hu10f11.x			
C 42	34.8	3.7	879	231	CNS04PDG	AL188205	Tetradon		
C 43	34.8	3.7	916	230	CNS01W26	AL169683	Tetradon		
C 44	34.6	3.7	558	204	AQ250056	T3C5-Sp6			
C 45	34.6	3.7	713	148	BF422416	Lr_ad_202			

ALIGNMENTS

RESULT 1
LOCUS CNS005TE/c
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR12K22 of RPL1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL060767
VERSION 1
KEYWORDS fruit fly.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 997)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPL1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp. the same strain used for the BDGP's p1 and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source location/Qualifiers

1. 997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPL1-98"
/clone="BACR12K22"

BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN

Query Match 5.1%; Score 48; DB 229; Length 997;

Best local Similarity 18.8%; Pred. No. 0.0034;

Matches 72; Conservative 124; Mismatches 187; Indels 0; Gaps 0;

QY	302	aataatccacacatatacgaacgctgagcttggaaacacgaatgctgtacattatg	361
DB	816	RAARAGRAAAR	757
QY	362	atggcaaaaagatgcccgaattattatggccacatgcaacaagaagaactaa	421
DB	756	ARARAGRAAAR	697
QY	422	ctacatgataagcgtggaatggaatggaatggaatggaatggaatggaatgga	481
DB	696	RRRRRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR	637
QY	482	ggagattttctatgccaagcgtactgctatgctgctgctgctgctgctgctgct	541
DB	636	GARRRRRRGRR	577
QY	542	tggagacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	601
DB	576	RGRRRRRRAR	517
QY	602	cagaagcgaagcgtgcgagctcattcgtgaaagaagcattgaatgataagctc	661
DB	516	RRRRRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR	457
QY	662	ctattccagaagcgtacagtt	684
DB	456	NN	434

RESULT 2
LOCUS CNS016E2/c
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106628
VERSION 1
KEYWORDS fruit fly.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1204)
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
source location/Qualifiers

1. 1204
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15A12"

Matches	86	Conservative	0	Mismatches	80	Indels	0	Gaps	0
QY	715	acgacatgatgtaatgcgctactcttcgglbgggaaatggaactatcaagatcaaga	774						
Db	788	ACTCAGATGAGAAAGCGCCTGCTTCACAGCACTGCCAAAGAGGTAGSCAAGACGACA	729						
QY	775	agcttaagcagcaaaaccattccttctatcagtgtygatitgaaggggggcgctatg	834						
Db	728	AGCTGAAAAAGTTCTCAGGCTGTTTTTTTNNAACTGTTTTTTGTAAAGTCTCGTGT	669						
QY	835	atctcggtagatgctctatcccttccaaaaaagagatcatatgt	880						
Db	668	ATTTCGTGTGTCGTTACTTCTCTTTTAAAAAGCTGTACTCTGT	623						

RESULT	5
A0675159/c	
LOCUS	A0675159 419 bp DNA 24-JUN-1999
DEFINITION	HS 2162, AI G09, F7C Cfr Approved Human Genomic Sperm Library D Homomaps genomic clone Plate=2162 COL=17 Row=M, DNA sequence.
ACCESSION	A0675159
VERSION	A0675159.1 GI:5207905
KEYWORDS	GSS.
SOURCE	human.

ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 419)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380389
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com)
BAC end Web server: <http://www.hnsc.washington.edu>
Plate: 2162 row: M column: 17
Seq primer: T7
Class: BAC ends
High quality sequence stop: 419.

FEATURES	source	Location/Qualifiers
1.	419	
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="P162 Col=17 Row=W"
		/clone.lib="C17 Approved Human Genomic Sperm Library D"
		/sex="male"
		/note="organ: sperm; Vector: pHELOBAC11; BAC clones in E-Coli DH10B"
BASE COUNT	110 a	99 c 77 g 133 t
ORIGIN		

Query Match	4.18;	Score 38.4;	DB 209;	length 419;
Best Local Similarity	46.38;	Pred. No. 1.8;		
Matches 126;	Conservative	0;	Mismatches 146;	Indels 0;
			Gaps	0.

[illegible]

QY	716	cgcaatgatgatgaatgcgcttactcttcgcgtrgggaaatggaactcaatcagatcgcga	775
Db	299	GGACAGTGCAGAAATCCACAGCTTTAGGGGGATCAAGATCATATTTGGCTAAATG	240
QY	776	gcttaagcagcaaaaacacatctcttccttatcagtgatgtaaggaggccgtatga	835
Db	239	TATTATATCTCCCAAGGTATGGGTGACAGTGAAGGTTGAAAGTGGGGGCGAGGGGTGAGG	180
QY	836	tctctgagatgattctatcccttcaaaaaa	867
Db	179	TACCACGTGATGTGTAATGATGACTTGCAATAATA	148

RESULT	6			
AA550191/c				
LOCUS	AA550191	395 bp	GSS	05-DEC-2000
DEFINITION	1307m3 gmbp:HB3.1, G. Roman Reddy Plasmidium		falciparum	genomic
	clone 1307m, DNA sequence.			

ACCESSION	AA550191
VERSION	AA550191.1
KEYWORDS	GI:2320443
SOURCE	GSS.
ORGANISM	malaria parasite <i>P. falciparum</i> .
REFERENCE	<i>Plasmodium falciparum</i>
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 395)
	Dame,J.B., Annot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z., Coppel,R.R., Comnan,A., Craig,A., Fischer,K., Foster,J., Goodman,N., Hinterberger,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A., Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su X.-Z., Thompson,J.K., Vital,F., Wellens,T.E. and Werner,E. Current status of the <i>Plasmodium falciparum</i> genome project Mol. Biochem. Parasitol. 79, 1-12 (1996)
TITLE	
JOURNAL	
MEDLINE	97001675
COMMENT	Contact: Dame JB

Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 339 4700
Tel: 352 339 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq Primer: T3
Class: shotgun
Location/Qualifiers

FEATURES	SOURCE	location/Qualifiers
		1..395
		/organism="plasmodium falciparum"
		/db_xref="taxon:5833"
		/clone="1307m"
		/clone_1b="gmbpfbH3.1, G. Roman Reddy"
		/lab_host="E. coli XL1-Blue"
		/note="Vector: Bluescript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Hondur HB3 isolate cultured in vitro, was digested with mung nuclease in the presence of 30% formamide at 500C (Veri, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6895). The ends of the fragment were ligated to EcoR V-cleaved and dephosphorylated bluescript SK(+). Recombinant plasmids transformed E. coli XL1-Blue."
BASE COUNT	135 a	49 c 35 g 175 t 1 others
ORIGIN		

Query Match	4.0%	Score 38.2	DB 201	Length 395
Best Local Similarity	49.5%	Pred No. 2.1		
Matches 97; Conservative	0	Mismatches 99	Indels 0	Gaps 0

QY 257 taatagctgaccagaactattctgttcagtgagcatcgcgaatgagatgacacactac 316
Db 389 TAATATACATTATGTTTCTTCTTAATAACAAATTCAGATGCAACAAGTTTTCGTGC 3300
QY 317 atgaaacggtgagccttgtaaaaaacagaatgctgtgtacattatgatgtgccaanaagatg 376

Db	329	AAGTGAATAATTTACAAAGAACAAGCAATCTTTAAACCTTTTAATAATTATNANTG	270
Oy	377	ccgagatattatttgcgccaatgatcacacaaagaagaactaccatcatgatgagc	436
Db	269	AATTATTAATATGTGGTCACAAATCAATTAAGATTAAGATTATGATATTTGAAAAATC	210
Oy	437	gtggagatggagatga	452
Db	209	ATGCACCAATTTGATGA	194
RESULT	7		
LOCUS	Bf628576		EST
DEFINITION	HVSMDB000618f Hordeum vulgare seedling shoot EST library		19-DEC-2000
ACCESSION	Bf628576		
VERSION	Bf628576.1	GI:11892734	
KEYWORDS	EST.		
SOURCE	barley.		
ORGANISM	Hordeum vulgare		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
AUTHORS	Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;		
	Hordeum.		
	1 (bases 1 to 712)		
TITLE	Development of a genetically and physically anchored EST resource		
JOURNAL	for barley genomics		
COMMENT	Unpublished (2000)		
	Contact: Wing RA		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Tel.: 864 656 7288		
	Fax: 864 656 4293		
	Email: twing@clemson.edu		
	Seq primer: AATTAACTCTACTAAGCG		
	High quality sequence start: 4		
	High quality sequence stop: 692.		
FEATURES	Location/Qualifiers		
SOURCE	1..712		
	/organism="Hordeum vulgare"		
	/cultivar="Morex"		
	/db_xref="taxon:4513"		
	/clone="HVSMDB000618f"		
	/clone_1ib="Hordeum vulgare seedling shoot EST library		
	HVCNN0002 (Dehydration stress)"		
	/tissue.type="Seedling shoot"		
	/lab_host="JUC121"		
	/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	245 a	137 c	176 g
ORIGIN		154 t	
Query Match	4.0%;	Score 38.2;	DB 151; Length 712;
Best Local Similarity	44.9%;	Pred.No.2.4;	Indels 0; Gaps 0;
Matches 145; Conservative	0;	Mismatches 178;	
Oy	500	caaacggtactgttcagtgcgattgaaaaggatcttgctttggagacgcagcagact	559
Db	75	CAAGTGAGCAAACCTCTTGATGGAAGAAGACATACAGAGATTTCAGAGCTCACGTAAAGC	134
Oy	560	cagacaacaacctacagattatatgatattatgcgccgaagaagatgcagaagcgcaagctgcgcg	619
Db	135	AAGATATACCTCTAAGCCAAACAGACACACAAAACCTAGTAGAAGATTCTCTAAGTGAAGC	194
Oy	620	agctctacgtaaaagagatgatgatgatgatgagctccgtcttatccagaacgcatata	679
Db	195	AGCAAGATCAGCAAAAGGCTAGCTGGAAGAACATATCCCGTACTACTTCTCTGGAGGTTGAAA	254

QY	680	casjtcacatgagacaaattgtagattgctctcaaacagacattgatgattcgcttact	739
Db	255	AGCGAGAGATGAGAGAGAAATTTGCTGTATTAAAGAAATTTCTCAGAGAGAGCTTCTG	314
QY	740	tttcgtgtgggaatlygaactatcaggatcagcaagcttaagcagcaanaaccattcc	799
Db	315	AAATTGCTGTACCAAGCAGCTTTCATGCAAGACAGAGAAACCTTGAAACACAGTTGC	374
QY	800	ttcttatcagtgatgattgaagg	822
Db	375	ATGATATGCGCTTGATGCTTGAG	397
RESULT	8		
LOCUS	A0137339	375 bp	DNA
DEFINITION	HS 3058.B1.E03.MR.CIT	Approved Human Genomic Sperm Library D	24-SEP-1998
ACCESSION	A0137339	sapiens genomic clone	Plate=3058 COL=5 ROW=J, DNA sequence.
VERSION	A0137339.1	GI:3528185	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 375) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A.	96 (17), 9739-9744	(1999)
MEDLINE	99380589		
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3058 row: J column: 5 Class: BAC ends High quality sequence stop: 375. Location/Qualifiers 1..375		
FEATURES	source		
	location		
	1..375		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="Plate=3058 COL=5 ROW=J"		
	/clone_11b="CIT Approved Human Genomic Sperm Library D"		
	/sex="male"		
	/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"		
BASE COUNT	121 a	71 c	67 g 115 t 1 others
ORIGIN			
	Query Match	4.0%	Score 38; DB 202; Length 375;
	Best Local Similarity	49.0%	Pred. No. 2.3; Indels 0; Gaps 0;
	Matches 101; Conservative	0; Mismatches 105;	
QY	739	tttcgtgtgggaatlygaactatcaggatcagcaagcttaagcagcaanaaccattcc	798
Db	55	TTTTGGTATAGTGTGGTAAATTATGAACAATCCCTAAATAAAGCAATTATAGATGCAATC	114
QY	799	cttcctacagtgatgattgaaggagccgatgatctctggtgagtgatgctatccct	858
Db	115	ATCTGAAGTGTCTTATGACAGAGCTTGTGGAAGATGAAGAGATCTTCATGATATTATG	174
QY	859	ttcaaaagagatcatatgtctgctccttaagctcttgagagatttaaacgcgaagga	918
Db	175	CTAATTTAAGGAAAAATCAGTTTCTGACTGTGATAGAGATGATCTCAATTAGGCAT	234

[illegible]

Db	373	GTGGCATTACTTAACGAGGATTTGAAGACAGACTCAATCATAGTTGTTATCAAAATGA	432
QY	591	ccgaagaagatgcagaaggaagcgtgcgcgagcttcattcctgaaaaagagccttgaagtgt	650
Db	433	ACAAAAATTAAGACCTTGTGTCTTGCTTGCTGCAGATGTAAGACAAATGTTACGAC	492
QY	651	agaagtcgcctcttcctccagagagcctac	681
Db	493	AGCGACCTTGTAGTATTCGACAGACTGATTGCA	523
RESULT	10		
LOCUS	AZ209683		
DEFINITION	AZ209683	889 bp DNA	GSS 31-AUG-2000
ACCESSION	AZ209683		
VERSION	AZ209683		
KEYWORDS	AZ209683.1 GI:8423656		
SOURCE	Strongylocentrotus purpuratus.		
ORGANISM	Strongylocentrotus purpuratus		
REFERENCE	Strongylocentrotus purpuratus		
AUTHORS	Euharipota, Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 889)		
TITLE	Cameron, R.A., Maharas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.		
JOURNAL	A sea urchin genome project: Sequence scan, virtual map, and additional resources		
MEDLINE	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)		
COMMENT	20402566 Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 138 row: A column: 3 Seq primer: T7 Class: BAC ends High quality sequence stop: 889. Location/Qualifiers 1..889 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone_lib="Plate=138 Col=3 Row=A" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BACe3.6; BAC clones in E-Coli DH10B"		
FEATURES	source		
BASE COUNT	267 a	188 c	134 g 300 t
ORIGIN			
Query Match	4.0%;	Score 37.6;	DB 217; Length 889;
Best Local Similarity	50.6%;	Pred. No.3.9;	
Matches	91; Conservative	0; Mismatches	89; Indels 0; Gaps 0;
QY	586	tatgacccgaaaagatgcagaagcgaagcgtgcgcgagcttcattcctgaaaaagagcattgaa	645
Db	163	TAGCTCTTAATGCACTTAATGCCATACACGCTGTGTAAGACAGACGTGATTTA	222
QY	646	gtgatagaggtccgcgtcattccagaagcgatcaggttcacacatgaaacaaattgagagt	705
Db	223	ATCCAACTCATGATGTGACTTTGCCAGATTACAAAAGATTATATATATATTAATTCAGTAA	282
QY	706	tcttctacacgacgatgatgatgatgcgtctactcttcgttggggaatgagactatga	765
Db	283	TTGACACACAAAGAAATTCGCTTATTCACGTAGTATATATGCTGTGATTTTGATTTACCA	342

RESULT	11
LOCUS	A0842470
DEFINITION	A0842470 799 bp DNA
ACCESSION	C661126A CPlOMAGDNA1 Cryptosporidium parvum genomic, DNA sequence.
VERSION	A0842470.1 GI:6003998
KEYWORDS	GSS.
ORGANISM	Cryptosporidium parvum.
SOURCE	Cryptosporidium parvum.
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
TITLE	Cryptosporidiidae; Cryptosporidium.
JOURNAL	1 (bases 1 to 799)
MEDLINE	Strong M.B. and Nelson R.G.
COMMENT	Preliminary profile of the Cryptosporidium parvum genome: an expressed sequence tag and genome survey sequence analysis Mol. Biochem. Parasitol. 107 (1), 1-32 (2000) 20183851
	Contact: Nelson, R. G.
	Dep't. of Medicine & Pharmaceutical Chemistry
	San Francisco General Hospital-University of California, San Francisco
	Box 0811, San Francisco, CA 94143-0811, USA
	Tel: 415 206 8846
	Fax: 415 206 3353
	Email: malaria@itsa.ucsf.edu
	For Annotation Data see http://medsfgh.ucsf.edu/jd/cprnags/home.html
	Seq primer: M13(-21) forward
	Class: shotgun.
FEATURES	Location/Qualifiers
source	1..799
	/organism="Cryptosporidium parvum"
	/strain="IOMA"
	/db_xref="taxon:5807"
	/clone_lib="CplOMAGDNA1"
	/lab_host="E. coli XL2 Blue MRF"
	/note="Vector: pBluescript II (SK-); Site:1; EcoRV: C. parvum (IOMA isolate) genomic DNA was hydrolytically sheared to produce fragments having a light size distribution between 2-4 kb by Dr. Ivonne Thorstenson of the Stanford DNA Sequencing and Technology Center (http://sequence-www.stanford.edu/group/techdev/shear.htm). The randomly sheared gDNA was chromatographed on Sephadryl S-400 to remove any small fragments and DNA eluting in the void volume was subcloned into an EcoR-vdigested, alkaline phosphatase-treated pBluescript II (SK-) vector and transformed into E. coli strain XL2 Blue MRF". Recombinant clones from the first plating of the library were selected for sequence analysis using T7 and T7 primers."
BASE COUNT	320 a 71 c 114 g 293 t 1 others
ORIGIN	
Query Match	3.9%; Score 37; DB 212; Length 799;
Best Local Similarity	51.5%; Pred. No. 5.7;
Matches	85; Conservative 0; Mismatches 80; Indels 0; Gaps 0.
Db	252 aaataagacgcgaccaggactatctgtcaagtcgatccgaatgatgaatacgcaa 311
Db	397 AATATTGGAATAATTAAATTGGAGTGTACTTTTATTCTATTAATAAANAATCAAGGTATCAA 456
Oy	312 catcatcagaacaacggcgtgagacttggaanaaacagaatgcgcgtgaccttatattgccaaaa 371
Db	457 AATTAACAACCAAAATTTTGAGGAAAATACTAACAAATTCAGGAATCTAATGAATTTAAAAAA 516
Oy	372 agatgcgagattattatcgcacacatgcacaaacagaagaaga 416
Db	517 AGATTGTGTAGTTATTGGCATTTTGAAGAANTAAATGAGAAAA 561
RESULT	12

[illegible]

CDNA, mRNA sequence.

ACCESSION BF727776
 VERSION BF727776.1 GI:12045637
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Eukaryote: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 411)

REFERENCE
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 1000052 row: F column: 04.
 Location/Qualifiers

FEATURES
 source
 1. .411
 /organism="Zea mays"
 /db_xref="dbEST:707007607.x2"
 /db_xref="taxon:4577"
 /clone_lib="1000 - Unigene 1 from Maize Genome Project"
 /note="This library represents the unique ESTs found in the first round of EST sequencing at Stanford University for the maize genome project. Sequences are present from libraries 486, 487, 496, 603, 605, 614, 618, 660, 683, 687, 707, and 945. Contigs were assembled using TRIM's CAP program and a representative EST from each contig was selected for the unigene set. All singlets were also selected."

BASE COUNT 100 a 73 c 85 g 153 t

ORIGIN
 1
 Query Match 3.9%; Score 36.8; DB 168; Length 411;
 Best Local Similarity 51.2%; Pred. No. 5.4;
 Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 565 acaacccacagatataatgataatgacgcaaaagatcagaagcgagcgagcact 624
 114 ACTGATCATCTTTCTTGACCATAGCTGGAAGATGAGTATTATGTTGGTTTGT 173

QY 625 catctgaaaagagcatgtagatgaggtccgcgtcattccagaagcgacatcagtt 684
 174 CTGGGATTTAAAGTACTATAGTGAAGTCTGTCGTCATTTCTTGAGTACATGTA 233

QY 685 caccatgaacaattgagatttgcttaacaagacatgtatgaatgc 732
 234 ATCCATTGACTCTTTTCAGAGTAATTCAGTGCATATGTTGATGC 281

RESULT 14
 CDS006DR 1101 bp DNA GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC: BACR3M19 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL072060
 VERSION AL072060.1 GI:4951799
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.

TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT The BDBG is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDBG_Drosophila_melanogaster_BAC_library was prepared by Kazuhiro Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDBG from the isogenic strain y2; cn bw sp, the same strain used for the BDBG's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 Location/Qualifiers

FEATURES
 source
 1. .1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPI-98"
 /clone="BACR3M19"
 /note="end : 17"

BASE COUNT 291 a 178 c 34 g 216 t 382 others

ORIGIN
 1
 Query Match 3.9%; Score 36.8; DB 229; Length 1101;
 Best Local Similarity 26.4%; Pred. No. 7.2;
 Matches 38; Conservative 54; Mismatches 52; Indels 0; Gaps 0;

QY 345 atgcgtgacattatgattgccaaaagatccgagatattatggccacaatgaac 404
 750 ATGTCCKKAWMTKAMARARMTAMAMADATRTAMAAATWDATKTTCTRSMAMC 809

QY 405 aacaagaagaaactactacacatgataagcgtgagaagatggatcttcgcgcg 464
 810 HGMAAATGKKKGAMTRACMGARAKAAAMKAMKATKKKAAADKATRKSKC 869

QY 465 tgaagaagaaagcgcgaggat 488
 870 TGAAMKKWMTAMRTKKRATDMRTX 893

RESULT 15
 CDS0039G 1101 bp DNA GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL063921
 VERSION AL063921.1 GI:4941778
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 JOURNAL Direct Submission
 COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
 The BDBG is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDBG_Drosophila_melanogaster_BAC_library was prepared by Kazuhiro Osoegawa and

OM of: US-09-462-846-2 to: N_Geneseq_0401.* out_format : pfs

Date: May 17, 2001 4:13 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frimex-p2n.model -DEV=xlip
-O=/cgn2_1/USPTO.spool/US09462846/runat_17052001_115429_970/app_query.fasta.1.376
-DB=N_Geneseq_0401 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPEXT=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MILEN=0 -MAXLEN=200000000
-USER=US09462846.@cgnl_1.198 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPX
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-462-846-2
Query length: 315
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890651
Search time (sec): 126.050000
```

score list:

Sequence	Strid	Orig	ZScore	EScore	len	Documentation
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25733 +			1692.00	3405.99	8.9e-182	945
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25735 +			1578.50	1953.57	3.3e-101	948
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25734 +			958.50	1919.06	5.9e-99	945
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X13110 +			874.00	1719.53	7.7e-88	12438
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X52181 -			841.00	1644.98	1.1e-83	25002
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:X74477 -			706.50	1382.79	4.4e-69	9623
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:X80389 -			685.00	1355.83	1.4e-67	2111
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:X74326 -			619.00	1215.71	8.9e-60	3759
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X20249 -			166.00	259.81	1.6e-06	116277
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X20248 +			166.00	257.26	2.8e-05	910715
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X5211 +			145.50	230.52	5.1e-06	6111
/SID6/gcgdata/geneseq/geneseqn/NA2001.DAT:X02175 +			145.50	243.64	1.2e-05	11461
/SID6/gcgdata/geneseq/geneseqn/NA2001.DAT:X02174 +			145.50	242.45	1.4e-05	12766
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X4213 +			145.50	242.30	1.5e-05	12949
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:X20087 +			145.00	249.53	5.8e-06	6103
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:X4332 +			115.00	203.04	0.0023	1650
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:X43143 +			109.00	194.53	0.0068	1182
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:X42063 +			101.50	98.86	1.4e-03	1830121
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:X74580 -			96.50	163.41	0.3656	2004
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:X0431 +			95.50	164.15	0.3323	1556
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X16518 +			93.50	151.50	1.68	3411
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X83879 +			93.50	151.50	1.68	3411
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X12108 +			93.00	158.12	0.7203	1699
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X16515 +			92.50	149.28	2.24	3471
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X83876 +			92.50	149.28	2.24	3471
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X23235 +			92.50	148.18	2.58	3837
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:X74512 +			92.00	141.72	5.91	6313
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X13047 +			89.50	128.24	33.25	21591
/SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:X97351 +			89.50	144.56	4.10	5482
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X09255 +			88.00	137.18	10.57	2897
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X50334 +			88.00	139.15	8.21	2897
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X47636 +			87.50	140.11	6.37	2896
/SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:X22989 +			87.50	141.63	6.81	3038
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X77722 +			87.50	127.63	35.99	9934
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X05585 +			87.00	151.58	1.67	1017
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X33226 +			87.00	147.13	2.95	1527
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X47470 +			87.00	145.71	3.54	1738
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X26232 -			87.00	141.69	5.92	2508
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X52337 +			86.50	133.93	16.04	5096
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X52153 +			86.50	117.22	136.66	21338

```
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X49283 + 86.00 147.91 2.67 1182
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X36468 + 86.00 146.36 3.25 1361
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X26612 + 86.00 142.54 5.31 1929
/SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:X37760 + 86.00 131.89 20.83 5100
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X13117 + 86.00 111.94 269.23 31517
```

```
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25733
```

seq_documentation_block:

```
ID X25733 standard; DNA; 945 BP.
XX
AC X25733;
XX
DT 21-MAY-1999 (first entry)
XX
DE B.subtilis cysteine protease Cpl coding sequence.
XX
KW Cysteine protease; Gram-positive bacteria; mutation; deletion;
XX proteolysis; expression; heterologous protein; ss.
XX
OS Bacillus subtilis.
XX
PN WO904016-A2.
XX
PD 28-JAN-1999.
XX
PE 14-JUL-1998; 98WO-US14529.
XX
PR 15-JUL-1997; 97EP-0305227.
XX
PA (GENV) GENENCOR INT BV.
XX (GENV) GENENCOR INT INC.
XX
PI Estell DA:
XX
WI: 1999-132260/11.
XX P-PSDB: W99371.
XX
PT Gram-positive bacterium with mutated or deleted gene for cysteine
PT protease 1, 2 or 3 - used to express proteins with reduced
PT proteolytic degradation, e.g. proteins, growth factors or enzymes
XX
PS Disclosure: Fig 1A-B; 31pp; English.
```

This sequence represents the Bacillus subtilis cysteine protease Cpl coding sequence. The invention relates to new Gram-positive bacteria, especially Bacillus, containing mutations in, or deletion of, at least part of the gene encoding cysteine protease (Cpl1, Cpl2 (X25734) or Cpl3 (X25735), so that CP-related proteolytic activity is lost. The bacteria are used for expressing heterologous proteins, particularly hormones, growth factors, cytokines or especially enzymes, e.g. protease, carboxylase, lipase, isomerases (tracemase, epimerase, tautomerase or mutase), transferase, kinase and phosphatase. Cpl-3 are used e.g. in soaps, dishwashing compositions, contact lens cleaners or laundry detergents, also for peptide hydrolysis, waste or textile treatment, to cleave fusion proteins and as animal feed additive.

Sequence 945 BP; 288 A; 180 C; 232 G; 245 T; 0 other;

alignment_scores:

```
Quality: 1692.00 Length: 315
Ratio: 5.371 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

```
US-09-462-846-2 x X25733 ..
```

```
Align seg 1/1 to: X25733 from: 1 to: 945
```

```
1 MetThrTrgIuProlEuphPhelysProValPhelysGluArgIleTr 17
|||||
1 ATGACGACTGAACCGTATTCTTCAAGCCTTTTCAAAAGAAATTTG 50
```

```

17 pglYgLYThrAlaLeuAlaaspPhegLYTYrThrIleProSerGlnArgT 34
   |||||
51 GGGGGGGACCCCTTACGTGATTTGGCTATACCATTCCTGCACACGAA 100
   |||||
34 hrGlyGluCysTrpAlaPheAlaAlaHISglnsngLYglnSerVal 50
   |||||
101 CAGGGGAGTGTGGCTTTGGCCCGCATCAAAATGGTCAAGCGTTGTT 150
   |||||
51 GlnsngLYmetTYrLYsgLYPheThrLeuSerGlnLeuProGlnHISh 67
   |||||
151 CAAGACGGAATGTATAGGGGTTCACCTCAGCAATATAGGGAACCTCA 200
   |||||
67 sArHISLeuPhegLYglnLeuGluGlyAspArgPheProLeuThrL 84
   |||||
201 CAGCATTTATTCGACAGCTTGAAAGGGGACCGTTCCCTGCTCTTACA 250
   |||||
84 ySIleLeuAspAlaAspGlnAspLeuSerValGlnValHISProAsn 100
   |||||
251 AAATATTAGATGCTGACACGACTTATCTGAGTGCAATCCGAATGAT 300
   |||||
101 GluTYrAlaAsnIleHISglnAsngLYglnLeuGlyLYsThrGluCys 117
   |||||
301 GAATATGCCAATACATCATGTAAGGCGTGACCTTGAAAAACAGAAATG 350
   |||||
117 pTYrIleIleAspCysGlnLYsAspAlaGluIleIleTYrGlyHISAs 134
   |||||
351 GTACATATTGATTTGCCAAAAAGATGCCGAGATTATTATGGCCACAA 400
   |||||
134 lArHrThrLYsgLYglnLeuThrThrMetIleGluArgGlyGluTYr 150
   |||||
401 CAACACAAAGAGAGAACTACTACCTAGATAGCGTGAGAAATGGGAT 450
   |||||
151 GluLeuLeuArgArgValLYsValLYsProGlyAspPhePheTYrVal 167
   |||||
451 GAGCTTGGCGCGGTGTAAAGCGGAGATTTTTCATATGTGCC 500
   |||||
167 oSerGlyThrValHISAlaIleGlyLYsgLYIleLeuAlaLeuGluTh 184
   |||||
501 AAGGGGACTGTTTCATGCGATGGAAGAAATTCCTTGTGAGACGC 550
   |||||
184 lncLInAsnSerAspThrThrArgLeuTYrAspTYrAspArgLYsAsp 200
   |||||
551 AGCAGAACTCAGACACACCTACAGATTAATATATATGACCAAAAGAT 600
   |||||
201 AlAGluGlyLYsLeuArgGluLeuHISLeuLYsSerIleGluValI 217
   |||||
601 GCAGAAAGCAAGCTGGCGAGCTTCATCTGAAAAAGACATGCAAGT 650
   |||||
217 eGluValProSerIleProGluArgHISThrValHISHisGluGlnI 234
   |||||
651 AGAGGTCCTCGTATTCACAGAACGCAATACAGTTCACCAAGCAAT 700
   |||||
234 lInAspLeuThrThrThrLeuIleGlyCysAlaTYrPheSerValGly 250
   |||||
701 AGGATTTGCTTACAGACACTTATGATGAGCGCTTACTTTCGGGGG 750
   |||||
251 LYsTrpAsnLeuSerGlySerAlaSerLeuLYsGlnLYsProPheLe 267
   |||||
751 AAATGGAACCTTATCAGATCAGCAACCTTAAAGCAACCAAAACCAT 800
   |||||
267 uLeuIleSerValIleGlyGluGlyArgMetIleSerGlyGluTYr 284
   |||||
801 TCTTATCAGTGTGATTTGAAGGGGAGCGGTATGATCTGTGGAGTAT 850
   |||||
284 aLYrProPheLYsLYsGlyAspHISMetLeuLeuProTYrGlyLeuGly 300
   |||||
851 TCTATCTCTTCAAAAAAGAGATCATATGCTGCTGCTTACGCTTGCA 900
   |||||
301 GluPheLYsLeuGluGlyTYrAlaGluCysIleValSerHISLeu 315
   |||||
901 GAATTTAACTCGAAGATATGACAGATATATGCTCCATCTG 945

```

```

seq_name: /SIDS6/gcdata/geneseq/geneseqn/NA1999.DAT:X25735
seq_documentation_block:
ID: X25735 standard; DNA; 948 BP.
XX
XX
AC: X25735;
XX
XX
DT: 21-MAY-1999 (first entry)
XX
DE: B. subtilis cysteine protease CP3 coding sequence.
XX
XX
KW: Cysteine protease; Gram-positive bacteria; mutation; deletion;
   proteolysis; expression; heterologous protein; ss.
XX
OS: Bacillus subtilis.
XX
PN: W09904016-A2.
XX
PD: 28-JAN-1999.
XX
PE: 14-JUL-1998; 98WO-US14529.
XX
PR: 15-JUL-1997; 97EP-0305227.
XX
PA: (GEMV ) GENENCOR INT BV.
   (GEMV ) GENENCOR INT INC.
XX
PI: Estell DA.
XX
DR: WPI: 1999-132260/11.
DR: P-PSDB: W99373.
XX
PT: Gram-positive bacterium with mutated or deleted gene for cysteine
   protease 1, 2 or 3 - used to express proteins with reduced
   proteolytic degradation, e.g. proteins, growth factors or enzymes
XX
PS: Disclosure; Fig 6A-B; 31pp; English.
XX
CC: This sequence represents the Bacillus subtilis cysteine protease CP3
   coding sequence. The invention relates to new gram-positive bacteria,
   especially Bacillus, containing mutations in, or deletion of, at least
   part of the gene encoding cysteine protease (CP1 (X25733), CP2 (X25734)
   or CP3), so that CP-related proteolytic activity is lost. The bacteria
   are used for expressing heterologous proteins, particularly hormones,
   growth factors, cytokines or especially enzymes, e.g. protease,
   carboxydase, lipase, isomerases (racemase, epimerase, tautomerase or
   mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
   CC soaps, dishwashing compositions, contact lens cleaners or laundry
   CC detergents, also for peptide hydrolysis, waste or textile treatment,
   CC to cleave fusion proteins and as animal feed additive.
XX
SQ: Sequence 948 BP; 286 A; 192 C; 244 G; 226 T; 0 other;
align_scores:
  Quality: 978.50      Length: 316
  Ratio: 3.822        Gaps: 1
  Percent Similarity: 81.013   Percent Identity: 56.646
alignment_block:
US-09-462-846-2 x X25735 ..
Align seg 1/1 to: X25735 from: 1 to: 948
1 MetThrGluProLeuPhePheLYsProValPheLYsgLuarGluIleTr 17
   |||||
1 ATGACGCAATACCGATTTTCTAAGCCCTGTGTAAAGAAAAAATCTG 50
   |||||
17 pglYgLYThrAlaLeuAlaasp...PhegLYTYrThrIleProSerGln 33
   |||||
51 GGGGGGAACCCCTTACGATAGATTTGGATACAGATTAATCTTCAAGAT 100
   |||||
33 rGThrGlyGluCysTrpAlaPheAlaAlaHISglnsngLYglnSerVal 49

```

```

101 CAACGGGGGAATGCTGGCCATTTCGGCTCATCCAAAGACCGAGCACT 150
50 ValGlnAsnGlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGlu 66
151 GTTGCAAAATGGCCCGGTATAAAGGAAGACATGATGATCGCTTGGGAGA 200
66 SHLSATGHTSLenPheGlyGlnLeuGluGlyAspArgPheProLeuLeu 83
201 GCACCGGGAAGTATTCGGCGCGGTAGAGGGGATCGGTTCCGCTCTGA 250
83 hrLysIleLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsn 99
251 CAAGAGCTCGGATGTGAAGAGATACGTCATTAAGTCCACCTGAT 300
100 AspGluTyrAlaAsnIleHisGlnAsnGlyGluLeuGlyLysThrGlu 116
301 GATTACTATGCGGAGAAACAGAGAGGAGAACTCGCAAGACGGAATG 350
116 STPTTyrIleIleAspCysGlnLysAspAlaGluIleIleTyrGlyHis 133
351 CTGGTACTATTCGACTGTAAAGAAAACGCAAAATCATTTACGGCAT 400
133 snalThrThrLysGluGluLeuThrThrMetIleGluArgGlyGluTrp 149
401 CGGCGCCCTCAAAAACCGAACTGTCAATGATCAACAGCGGTGACTG 450
150 AspGluLeuLeuArgValLysValLysProGlyAspPheThrVal 166
451 GAGGCGCTGCGGAGAAATCAAAATTAACCGGTGATTTCTATATG 500
166 LProserGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGlu 183
501 GCCAGGGAACGCTGCACGATGTGCAGAGGGGCCCTTGTATGAGA 550
183 hrGlnAsnSerAspThrThrTyrArgLeuTyrAspTyrAspArgLys 199
551 CTCGCAAAATTCGATGCCACATACCGGGTGTACGATTTGACCGCT 600
200 AspAlaGluGlyLysLeuArgGluLeuHisLeuLysSerIleGluVal 216
601 GATAGCAACGGAAGTCCGAGAGACTTCATTTGCCAAAGCGGTCAATG 650
216 ILeGluValProSerIleProGluArgHisThrValHisGlnGln 233
651 CGCCAGGTTCCCATGTGGACGGGTATATAGATGAATCGACGAATCA 700
233 legIuAspLeuLeuThrThrLeuIleGluLysAlaTyrPheSerVal 249
701 GAAGAAGATATACCATTAATAACATTTGTCAAGGGAATTTTTCGCT 750
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysPro 266
751 TATAAATGGGACATCAATGGGGAAGCTGAAATGGCTCAGAGTGAATC 800
266 eleuLeuIleSerValIleGluGlyGluGlyArgMetIleSerGlyL 283
801 TCTGATTTGCGAGCTGATAGAGAGACGGTTGCTCAAGTATAGAGACA 850
283 yValTyrProPheLysGlyLysAspHisMetLeuLeuProTyrGlyLeu 299
851 AAACATGTCGCTCAAAAAGGTGATCATTTATTTGCGCGCTCAATG 900
300 GlyLysPheLysLeuGlyLysTyrAlaGluLysIleValSerHisLeu 315
901 CCCGATTTTACGATAAAGAACTGTATCCCTATCGTGTCTCATAT 948
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA199.DAT.X25734
seq_documentation_block:
ID X25734 standard; DNA; 945 BP.
XX
AC X25734;

```

```

XX
XX 21-MAY-1999 (first entry)
DE B. subtilis cysteine protease CP2 coding sequence.
XX
XX Cysteine protease; Gram-positive bacteria; mutation; deletion;
XX KM proteolysis; expression; heterologous protein; ss.
XX OS Bacillus subtilis.
XX
XX W09904016-A2.
XX
XX 28-JAN-1999.
XX
XX 14-JUL-1998; 98WO-US14529.
XX
XX 15-JUL-1997; 97EP-0305227.
XX
XX (GEMV ) GENENCOR INT BV.
XX PA (GEMV ) GENENCOR INT INC.
XX
XX Estell DA:
XX
XX WPI: 1999-132260/11.
XX DR P-PSDB: W99372.
XX
XX Gram-positive bacterium with mutated or deleted gene for cysteine
XX PT protease 1, 2 or 3 - used to express proteins with reduced
XX PT proteolytic degradation, e.g. proteins, growth factors or enzymes
XX
XX Disclosure; Fig 5A-B; 31pp; English.
XX
XX This sequence represents the Bacillus subtilis cysteine protease CP2
XX coding sequence. The invention relates to new Gram-positive bacteria,
XX especially Bacillus, containing mutations in, or deletion of, at least
XX part of the gene encoding cysteine protease (CP1 (X25733), CP2 or CP3
XX (X25735), so that CP-related proteolytic activity is lost. The bacteria
XX are used for expressing heterologous proteins, particularly hormones,
XX growth factors, cytokines or especially enzymes, e.g. protease,
XX carboxylase, lipase, isomerases (racemase, epimerase, tautomerase or
XX mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
XX soaps, dishwashing compositions, contact lens cleaners or laundry
XX detergents, also for peptide hydrolysis, waste or textile treatment,
XX to cleave fusion proteins and as animal feed additive.
XX
XX Sequence 945 BP; 290 A; 189 C; 223 G; 243 T; 0 other:
XX

```

```

alignment_scores:
  quality: 958.50      length: 313
  ratio: 3.804        gaps: 1
  Percent Similarity: 80.511  Percent Identity: 55.591

```

```

alignment_block:
US-09-462-846-2 x X25734 ..

```

```

Align seg 1/1 to: X25734 from: 1 to: 945

```

```

3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTrpGlyL 19
4 ACCCATCCATTAATTTTAAAGCCTGTCTTAAAGAAAGCTATGGGAGG 53
19 yThrAlaLeuAlaAsp...PheGlyTyrThrIleProSerGlnArgThr 35
54 GACGAGCTTCGTGACGCTTTGGCTACGCATACCTTCAACAAAACAG 103
35 LysGlnSerTrpAlaPheAlaAlaHisGlnAsnGlyGlnSerValValG 51
104 GTGAGTGTGGCGCTTCTGCAATGCCATGAGCTCGTGTGTAAAA 153
52 AsnGlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGluHisArg 68
154 AATGGCCCGCTGCGAGAAAGACATTTGATCAAGTATGGAAGATCATCC 203

```

```

KW vaccine attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
PN WO9805055-A2.
XX
XX 12-NOV-1998.
PD
XX
XX 04-MAY-1998; 98WO-US08985.
PF
XX
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046555.
XX
PA (HUMA-) HUMAN GENOME SCT INC.
PI Barash SC, Dillon PU, Kunsch CA;
PI WPI: 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 963-969; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SO Sequence 12438 BP; 4071 A; 2218 C; 2424 G; 3708 T; 17 other;

alignment_scores:
      Quality: 874.00          Length: 317
      Ratio: 3.553             Gaps: 3
Percent Similarity: 77.603     Percent Identity: 53.628

alignment_block:
US-09-462-846-2 x X13110 ..

Align seg 1/1. to: X13110 from: 1 to: 12438

   4 GIUPTOLEUPHEPHELSPROYALPHELYSGUARGLIETTPGILGYLTHN 20
      |||||||.....:|||||.....:|||||.....:|||||.....
6276 GAMCATTATTTTACACCTGTTCAGCAAAAAAATTGGCGGCACAA 6325
      20 FALDENALAASP...PheGLYTYrThrLIleProSerGIaArgHcgly 36
         |||  |||  |||.....:|||||.....:|||||  |||
6326 TCgcTCTAcAcCcCTATtTGtTGtTCgATTCcCGAcGcATMAAtTtgGtG 6375
      36 IucSTrPALpheAlaAIahISglAnaNglyGINserValalGlAsn 52
         ||  |||||.....:|||||  :|||  |||||.....|||
6376 AGAgTtGGcCATCACTGCACATCCACATGCGGTAgtACTGTTTTAAAT 6425
      53 GlyMetYLysGLyPherThIeuSeRgLLeutRpLUhtSHIsArghI 69
         |||  ::::| |||||  |||:::||||| |||  |||:::
6426 GGcGAATTTMAAGAAAATAATTAATGAATTTGGGcAGAcATCAAGA 6475
      69 slenPhgLIgInLeUGLIYAAsPaRgPhPrOleuThrLySLel 86
         :|||:::|||||  |||

```

```

6476 GTTATTGGCCATGAGGAGGCGCTGCTTCATTATTAAAGAAATTT 6525
86 euAspalaAspGlnAspLeuSerValGlnValHisProAsnAspGluTyr 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6526 TAGAGCGCTGAAGATGATTTATCCGTGCAAGTTCATCGATGATGCATAT 6575
103 AlaAsnIleHisGluAsnGluGluLeuGlyLeuThrGluCysTrpTyrIle 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6576 GGCCAAAGACATGAA...GGCGAGCTAGGCAAAAGTGAATGTTGGTACAT 6622
119 eileAspCysGlnIysAspAlaGluIleIleTyrGlyHisAsnAlaThrT 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6623 TATTGATGAGAAACCTGGCGCTGAATCATTTATGACATCAGCGAAAA 6672
136 hrlYsGluGluLeuThrThrMetIleGluArgGlyGluTrpAspGluLeu 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6673 CAAAGAAAGAACTGCGAGAAAGATGAAAGAGTGGTGGGACATCTT 6722
153 LeuArgArgValIysValIysProGlyAspPhePheTyrValProSerGlu 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6723 TTAAAGAAAGTACCAAGTAAAGGCGATTTTATATGATGCCAAGCGG 6772
169 yThrValHisAlaIleGlyIysGlyIleLeuAlaLeuGluThrGlnGln 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6773 CACTATTTCATGGCATGGCTCGGTATTTTATTTAGAAACGCAACAA 6822
186 snSerAspThrThrThrTyrArgLeuTyrAspTyrAspArgLysAspAlaGlu 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6823 GGAGTGTATCTACGTATCGGTTTATGATTATGATGATCAACATGATCAA 6872
203 GlyIysLeuArgGluLeuHisIleLysSerIleGluValIleGluVal 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6873 GGAAGAACCCGGGAATGTCATATTCACAAATCCATTTGATGATCAACACCT 6922
219 lProSerIleProGluArgHisThrValHisIleGluGlnIleGluAspL 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6923 TCCGGGAGAGGATCCGACCTCTATTCACACAGAAATCAAGACAT 6972
236 euleuThrThrThrIleuIleGluCysAlaTyrPheSerValGlyLysTrp 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6973 CAAGCTTGTCTTATGTGTAAGAAACAGATTCTTTAATCTATGATGATGG 7022
253 AsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeuIle 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7023 CGTGTAGAGGGATTTCTAAAGTCAAAACAAAGCACCCTTATACATTAC 7072
269 eSerValIleGluGlyGluGlyArgMetIleSerGlyGlu..... 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7073 AACAGTATTTGAAGGTGTGGTGTGATTGATTACAGAAAGATCTGCAAAAG 7122
283 .....TyrValTyrProPheLysGlyAspHisMetLeuLeuPro 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7123 CAGAGTGGCTACTTTGATTTGAAAAAGGCGATAGCTTTATCTTTCCG 7172
297 TyrGlyLeuGlyGluPheLysLeuGluGlyTyrAlaGluCysIleValSe 313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7173 ACAGATATTCGAGCTGGCGTTTGAAGTGAATCAAGCATATATAGCTTC 7222
313 r 313
7223 A 7223

```

seq_name: /SID6/gcdata/geneseq/geneseqn/NA1998.DAT:V52181

seq_documentation_block:

ID V52181 standard; DNA: 25002 BP.

AC V52181;

XX 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:48.

DE

```

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM computer readable medium; vaccine; pharmaceutical composition; ds.
XX Streptococcus pneumoniae.
XX WO9818931-A2.
XX 07-MAY-1998.
XX 30-OCT-1997; 97WO-US19588.
XX 31-OCT-1996; 96US-0029960.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
XX Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
XX
XX PS Claim 1; Page 431-445; 1409pp; English.
XX
XX CC The present invention describes a computer readable medium which has
XX CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX CC on it, or a representative fragment or a sequence at least 95% identical
XX CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX CC pneumoniae. The present invention also describes an isolated nucleic acid
XX CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX CC by a process comprising: (a) screening a genomic DNA library using as a
XX CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX CC to 391, identifying members of the library which contain sequences
XX CC that hybridize to the target sequence and isolating the nucleic acid
XX CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX CC from an organism, amplifying nucleic acid molecules whose nucleotide
XX CC sequence is homologous to amplification primers derived from the
XX CC fragment of the S. pneumoniae genome to prime the amplification and
XX CC isolating the amplified sequences. The computer readable medium can be
XX CC used in a computer-based system for identifying fragments of the
XX CC S. pneumoniae genome of commercial importance, or expression modulating
XX CC fragments of the S. pneumoniae genome. Products from the present
XX CC invention can be used in diagnosis kits and assays, and pharmaceutical
XX CC compositions and vaccines for S. pneumoniae.
XX
XX SQ Sequence 25002 BP; 7422 A; 5498 C; 4618 G; 7463 T; 1 other;

```

alignment_scores:

Quality: 841.00 Length: 316

Ratio: 3.461 Gaps: 5

Percent Similarity: 76.899 Percent Identity: 53.797

alignment_block:

US-09-462-846-2 x V52181/rev ..

Align seg 1/1 to reverse of: V52181 from: 1 to: 25002

```

3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTrpGly 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2381 TCAGAACCATTTATTTTACAAATCAAGTATGCAAGAAAAAATCTGGGCTGG 2332
19 yThrAlaLeuAlaAsp...PheGlyTyrThrIleProSerGlnArgThr 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2331 AGCCAAAGCTACGTATGATGTTGGCTAGCAGATCCCAAGTAAAAAAATCG 2282
35 yGluCysTrpAlaPheAlaHisGlnAsnGlyGlnSerValValGln 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2281 GAGATATTGGCCATCTCAGCCCATCCAAATGGAGTCTAAAGTTGCC 2232
52 AsnGlyMetTyrGlyPheThrLeuSerGluLeuTrpGluHisHisIar 68
||||| ||||| ||||| ||||| |||||
2231 AATGTCCTTACGAGGGAACAGATCTTGTCTACTTGTATGGGGAACACCG 2182
68 gHisLeuPheGluGlnLeuGluGlyAspArgPheProLeuLeuThrIysI 85
||||| ||||| ||||| ||||| |||||
2181 TGAATTATTGGCAATCTCCAGAACCTGTATTTCACATTTCAGCCACAGA 2132
85 leLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 101
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2131 TCCTGATGCCACGACACTGCTCAGTCTCCAGTTCCACCCAGACGATGCT 2082
102 TyrAlaAsnIleHisGluAsnGlyLeuLeuGlyLysThrGluCysTrpTy 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2081 TATGACTCGAGCATGAA...GGCGAACTCGCAAAACAGAAATGCTGTGTA 2035
118 rIleIleAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsnAlar 135
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2034 CATATCGACGCGGATGAAAGTTCAGAGATTATCTACGGTCACAAATCCA 1985
135 hTrpLysGluGluLeuThrThrMetIleGluAspGlyGluTrpAspGlu 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1984 AGTCAAAAGAAAGAACTCCGCCAGCAATCCAGGACAGAAAGTGGATGAC 1935
152 LeuLeuArgArgValLysValLysProGlyAspPhePheTyrAlaProse 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1934 TTGTTGACCAAAAGTCCCTGTTAAGCTGAGAGATTCTTATATGATACCA 1885
168 rGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGln 185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1884 CGGAGCTATGACGCTATCGGGGGGTATCTTAATCTTGAAACCACGAC 1835
185 lnaAsnSerAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspAla 201
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1834 AGTCTAGCAGATCCACCTATGCGCTATGACTTTCACCCGACAGAGATGAC 1785
202 GluGlyLysLeuArgGluLeuHisLeuLysSerIleGluValIleG 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1784 AAAGCAAACTTGCCTGAACTCATCTTGAAAAATCATCATGATGCTTGA 1735
218 uVal.....ProSerIleProGluArgHisThrValHisHisGluGln 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1734 CATTTGGTGAAGCCAGCAAAATAGCCGCTCTGTAACGTGT.....AAAG 1694
233 leGluAspLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerVal 249
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1693 CAGATGATTTGCGTTCACCTCTCTATCTATGATGATTTCTTCGCGAGTT 1644
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPh 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1643 TACAAGTGGGAAATTAAGTGAAGTTGAAAGAAAGACACCTGACATA 1594
266 eleuLeuIleSerValIleGluGluGlyArgMetIleSerGlyGlu 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1593 CAGCTTTTGTGAGTCTTGGCTGCTCAGGTCAGCTAAGTGTGACGGGA 1544
283 yValTyrTrpPheLysGlyAspHisMetLeuLeuProTyrGlyLeu 299
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1543 AAAATATTCCAATTCAAAGGACGACATTATCTACCAAGATGATGTT 1494
300 GlyLysLeuLeuGluGlyTyrAla..GluCysIleValSerHis 314
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1493 GAACTTGGACTCTGGAAGGCAAGTTTGAAATGATTTGTAGCAT 1446
seq_name: /sids6/genegdata/geneseq/NA1997.DAT:V74477
seq_documentation_block:
ID V74477 standard: DNA, 9623 BP.
AC V74477:
XX

```

```

DT 16-MAR-1999 (first entry)
XX
XX Staphylococcus aureus contig SEQ ID #166.
DE
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
OS
XX
XX Key Location/Qualifiers
FH 361..420
FT misc_feature
FT /tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT 2161..2220
FT misc_feature
FT /tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT 3961..4020
FT misc_feature
FT /tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT 5761..5820
FT misc_feature
FT /tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT 7561..7620
FT misc_feature
FT /tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT 9361..9420
FT misc_feature
FT /tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
EP786519-A2.
XX
XX 30-JUL-1997.
XX
XX 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA.
XX
XX PI Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -
XX stored on computer readable medium and used in the production of
XX anti-S.aureus vaccines
XX
XX Claim 1; Page 816-822; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer readable
XX medium, preferably selected from a floppy or hard disk, random access
XX

```


0945 CACGCGCCGGCAGGCTCGGTATTTTAAATTGAGGCACCACCAATAA 37

+

that bind and inhibit the activity of the polypeptide. Such compounds

alignment_scores:

Quality: 619.00 Length: 299
Ratio: 2.906 Gaps: 4
Percent Similarity: 71.237 Percent Identity: 39.465

alignment_block:

US-09-462-846-2 x V74326/rev ..

Align seg 1/1 to reverse of: V74326 from: 1 to: 3759

```

3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTyrPheLys 19
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
972 ACAATGCCATTTATTTTACACCAATTTTAAACAAATTTATGGGGCGG 923
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19 yThrAlaLeuAlaAspPheGlyTyrThrIleProSerGlnArgTyr.Gly 35
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
922 TCACGCTCTAAGTGGATTGATTCATTAAGCAATGATACACTGGGG 873
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
36 GluCysTrpAlaPheAlaAlaHisGlnAsnGlyGlnSerValValGlnAs 52
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
872 GAATGTTGGGTGTCTCACACATCCAAATGGTACGAGCGAGATTATTAA 823
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 nglyMetTyrLysGlyPheThrLeuSerGluLeuTyrGlnHisArgH 69
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
822 TGGACCATATCAAGTGCACAAATTAAGCCGTATTGTCAGAACATCGTG 773
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 IsteuPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrIste 85
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
772 AATTGTTGGGTGTCTCCACAGCAAGATTTCGCTTCACTAAATAA 723
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 LeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 102
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
722 GTGGATGCAAGAGATCTTCTATTCAATGTCACACCTGATTAATCTTA 673
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 rAlaAsnIleHisGlnAsnGlyLeuGlyLysThrGluCysTyrPtyr 119
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
672 TCGTATTGAGCATGAACGCGGCAATATGCAAAATCTGAATGTGTATA 623
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 IetIleAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsnAlaTr 135
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
622 TTATATGATGCAAGAGATGCAAGAAATAGTATAGGCAATTAGACAG 573
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 ThrLysGluGluLeuThrMetIleGluArgGluTyrAspGlu 152
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
572 TCTAGAGAAAGAGTTCGATCTGTTCAACAGGACATAGATCGAT 523
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 uLeuArgArgValLysValLysProGlyAspPhePheTyrValProSer 169
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 ACTTGGATATATTAAGTAAACCTGCGAATCTATTATNNNNNNNN 473
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 LyrThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGln 185
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 423
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
186 AsnSerAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspAla 202
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 NNTCTGAGACATTTACATATGACTTTTGTATTCATCGTCAGATTAAT 373
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 uGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleGlu 218
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 ATATATGATGATGACGCTTAATATTTGAAAAAGCTTATAGCGTTAT 323
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 .....ValProSerIle..ProGluArgHisThrValHisHis 230
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 ACAATGCCATTTACTTAATATTTTCCGCGAA.....AGC 288
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 GluGlnIleGluAspLeuLeuThrThrThrLeuIleGluCysAlaTyrPh 247
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 GAATATTATGAAACCATTAAGTGTACACACATTTGTATCGAATGATTCT 238
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
247 eSerValGlyLysTyrPsnLeuSerGlySerAlaSerLeuLysGlnGln 264
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 TACATTGGTTAAATGGCAATTTCTGGACAGCTAAATTAATATGAAAGCTA 188
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 ysrProPheLeuLeuIleSerValIleGluGlyGluArgMetIleSer 280
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
187 GAGACTTCTGTAGTTAGTACAGTGTGGAAGCGCAAGGCAATATGATGTC 138
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 GlyTyrValTyrProPheLysLysGlyAspHisMetLeu 295
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
137 GATGTTGAATTTTCAAACTGACTACTGTTACAAACTTTATTTTG 93
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.X20249

seq_documentation_block:

ID X20249 standard; DNA; 116277 Bp.

XX X20249;

DT 04-MAY-1999 (first entry)

XX Borrelia burgdorferi polynucleotide sequence #2.

KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

XX infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

PN W09858943-A1.

XX 30-DEC-1998.

PF 18-JUN-1998; 98MO-US12764.

XX 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (MED1-) MEDIMUNE INC.

PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

PI White OR;

DR WPI; 1999-081217/07.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop

PT products for the detection, diagnosis, characterisation, prevention

PT and therapy of infections, particularly Lyme disease

XX Claim 1; Page 672-737; 1128pp; English.

XX X20248 to X20402 represent polynucleotide sequences isolated from

CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for

CC the detection, diagnosis, characterisation, prevention and therapy of

CC BB infections, e.g. Lyme disease. They can also be used for the

CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.

CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and

CC endemic relapsing fever, and Lyme borreliosis, more commonly known as

XX Lyme disease.

XX Sequence 116277 Bp; 42656 A; 19868 C; 14490 G; 39250 T; 13 other;

alignment_scores:

Quality: 166.00 Length: 406
Ratio: 0.883 Gaps: 22
Percent Similarity: 46.305 Percent Identity: 23.645

alignment_block:

US-09-462-846-2 x X20249/rev ..

```

63094 GAGTTGGTCATCTTTGGAGCAAGCAAGTATCATCTTTTAATCCCATTTT 63045
211 LvsLysSerIleGluValIleGluValProSer..... 221
:::|||||:::|||||:::|||||:::
63044 CAAGATATATTAGCGATTATTAGACTCCAAATACTAATTTGAATGTGAT 62995
222 .....IleProGluArgHisThrValHisHisGluGlnIleGluA 235
|||||:::|||||:::
235 sPLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyIys 251
|||||:::
62994 TCTTCTGATGTTTAAT.....GGTGC 62972
252 TrpAsnLeuSerGlySerAlaSerLeuLysGluGlnLysProPheLeu 268
:::|||||:::|||||:::
62921 GTCAGCTATAAATAAATCCTTAAATCTTAAGAAA..... 62889
268 uLeSerValIleGluGluGluGlyArgMetIleSerGlyGluTyrValT 285
|||||:::
62888 .....GGTGAGACCAAT 62876
285 yTrpProPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyIu 301
:::|||||:::|||||:::
62875 TTATAGCTAAAAAACGACAAAACCTGTTTATTGAT...GGGAGCGCGCA 62829
302 PheLysLeuGluGluTyr 307
:::|||||:::
62828 GCTTTTATTGCTGGTTTT 62811

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.x20248
seq_documentation_block:
ID X20248 standard; DNA; 910715 bp.
XX
AC X20248;
XX
DT 04-MAY-1999 (first entry)
XX
DE Borrelia burgdorferi polynucleotide sequence #1.
XX
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KN infection; diagnosis; characterisation; detection; ds.
XX
OS Borrelia burgdorferi.
XX
PN MO9858943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98MO-US12764.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDT-) MEDIMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO;
PI White OR;
XX
DR WPI: 1999-081217/07.
XX
PR New isolated Borrelia burgdorferi nucleic acids - used to develop
PR products for the detection, diagnosis, characterisation, prevention
PR and therapy of infections, particularly Lyme disease
XX
PS Claim 1; Page 157-671; 1128pp; English.
XX
CC X20248 to X20402 represent polynucleotide sequences isolated from

```

CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.

XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

alignment_scores:

Quality: 166.00 Length: 406
 Ratio: 0.863 Gaps: 22
 Percent Similarity: 46.305 Percent Identity: 23.645

alignment_block:

US-09-462-846-2 x X20248 ..

Align seg 1/1 to: X20248 from: 1 to: 910715

```

1 MetThrThrGluProLeuPhePhe.....LysProValPheLeuGluLys 15
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
491204 ATGAATATGACATATATATTTTAAATGAAAAAATAATTAAGAATA 491253
15 gllETPRGlyGlyThrAlaLeuAlaSP.....PheGlyTyrThrIleP 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491254 TGATTTGGGGGGGGAATTAATTTTATCCCAATCTTTGGGATAGATTTG 491303
30 roserGlnArgThrGlyGluCysThrAlaPheAlaIleHisGlnAsnGly 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491304 ATGGAAAGCCCAAGCGTGAATGTGG...CTTGAGCAGCAGACAGACATT 491350
47 GlnSerValValGlnAsnGlyMetTyrLysGly.....PheThrLeuSe 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491351 TCCTAGTAAGATT.....TGTATATAAAATGAATATGCTCTTTAAG 491391
61 rGluLeuTrpGluHisHisArgHisLeuPheGlyGlnLeuGluGlyAsp 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491392 CGATTTTATAGAAAGCATTAAGAGCTTTAGGCTGT.....AAAGCAG 491435
78 rGpPheProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeuSerVal 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491436 AATTTCCTTTTTCCTTAGGATATGTCTGCAATAAAGCCCGTGTGATT 491485
95 GlnValHisProAsnAspGlu.....TyrAlaAsnIleH 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491486 CAAATTCATCCTCTTAAGATATTGCCTTAAGAGGTATGATCAGAGAA 491535
106 sGluAsnGly.....GluL 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491536 TAAATPAAAGGGATAGACATTAATGATCCCAAAAGACATACAAAGACAAA 491585
111 euGlyLysThrGluCysTrpTyrIleIle..... 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491586 ACCCCAAAATTGAATATTATTCCTTAGTGATTTTATGCTTAATA 491635
121 .....AspCysGlnLysAspAlaGluIleIle..... 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491636 GCGTTTATACCTTAGATGAGATTAAAAAATTTATGAAATTCGGAATT 491685
130 .....TyrGlyHisAsnAlaThrThrLysGlu..... 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491686 AAATTTCAGCTTTCATACATTAAGATTTTGTAAAGACATTTTGTGATT 491735
139 .....GluLeuThrMetIleGlu..... 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491736 TACAAATGTATGAACTGAGAAAGATTATTTGAAAAAATTTTAAAAAATTG 491785
146 .....ArgGlyGluTrp...AspGluLeu..... 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491786 GATCTATTGATTAATTTTATGAGGCTATTGGTTTAAATGAATTTTACATAT 491835

```

```

153 .....LeuArgArgVal 157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491836 TTATGGTATAGATGTGGCCCTTTGGTATTTTATAGATATGAATTTTAA 491885
157 ysValLysProGlyAspPhePheTyrValProSerGlyThrValHisAla 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491886 AACATAAACAGAGAAAGTCTTATACAAATAGTACGAGAGTCAAGCA 491935
174 IleGlyLysGlyIleLeuAlaLeuGluThrGlnGlnAsnSerAspThr 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491936 TATCTTAAGGAGAT...TGCATTGAGCTTATGACCAATTCGCAATGT 491982
190 TYArg.....LeuTyrAspTyrAspArgLysAsp..... 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491983 TATTAGGCGCTGGCGCTTACTACAAAGTATATGATTAAGACAGATGTTAA 492032
201 .....AlaGluGlyLysLeuArgGluLeuHis.....Leu 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492033 GAGTTGTCATATTGAGGAGAAAGTATATCATTTTAAATCCGATTTT 492082
211 LysLysSerIleGluValIleGluValProSer..... 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492083 CAAGATTAATTTTAGCGGTATTTAGACTCCAAATCTAAATTGAAATTGAT 492132
222 .....IleProGluArgHisThrValHisIleGluGlnIleGlu 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492133 TCAAAAAATAAATAATGAGAACATTGTATTAATGAATAGTGCATAG 492182
235 sPheLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLys 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492183 TCCTCTAGTTTAAAT.....GGGTGC 492205
252 TrpAsnLeuSerGlySerAlaSerLeuLysGlnLysProPheLeu 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492206 GTGAGATPAAATAAATCCTTAATCTTAAGAA..... 492238
268 uIleSerValIleGluGlyGlyArgMetIleSerGlyGlyVal 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492239 .....GGTGAACCATAT 492251
285 yTrpPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGlu 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492252 TTATAGCTAAAAAGCAGAAACTGTTTATGAT...CGGACGCGGAA 492298
302 PheLysLeuGluGlyTyr 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492299 GCTTTATTGCTGGTTT 492316
seq_name: /SID56/gcdata/geneseq/geneseqn/NN2000.DAT.A54211
seq_documentation_block:
ID A54211 standard; DNA; 6111 BP.
XX
XX A54211;
XX
XX 26-FEB-2001 (first entry)
XX
XX DE Transformation vector pCIB9818.
XX
XX Tricothecene resistance; resistant; crop protection; mycotoxin;
XX fungus; wheat; maize; barley; rice; heterologous gene;
XX transformation; Fusarium; ds.
XX
XX Synthetic.
XX
XX WO2000060061-A2.
XX
XX 12-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-EP02769.
XX
XX 31-MAR-1999; 99US-0282995.

```



```

XX plasmid pZU634 for Potato virus Y (PVY) resistance in tomato.
DE
XX Potyvirus: PVY: virus resistant; virus tolerant;
XX tomato; coat protein; CP; plasmid pZU634; ds.
XX
XX Chimeric - Potato virus Y.
OS Chimeric - Bacteria.
OS Chimeric - Agrobacterium sp.
OS Chimeric - Unidentified.
XX
XX Location/Qualifiers
FH Key 55..1786
FT /tag= a
FT /note= "Ubiquitin3 (Ubi3) promoter plus leader
FT intron"
FT 1790..3847
FT /tag= b
FT /note= "Potato virus Y (PVY) inverted repeat region"
FT 1790..2574
FT /tag= c
FT /rpt_type= INVERTED
FT /note= "Corresponds to PVY coat protein (CP) region"
FT 2595..3036
FT /tag= d
FT /label= Actin2_intron1
FT complement (3057..3847)
FT /tag= e
FT /rpt_type= INVERTED
FT /note= "Corresponds to PVY CP region"
FT 3855..4124
FT /tag= f
FT /label= NOS
FT /note= "Agrobacterium nopaline synthase terminator"
FT 4193..5364
FT /tag= g
FT /label= SMAS_promoter
FT 5386..6561
FT /tag= h
FT /product= "Phosphomannose isomerase A"
FT 6623..6897
FT /tag= i
FT /label= NOS
FT /note= "Agrobacterium nopaline synthase terminator"
FT 6949..7080
FT /tag= j
FT /label= Nopaline_left_border_fragment
FT 7359..8147
FT /tag= k
FT /product= "Bacterial spectinomycin-detoxifying enzyme,
FT aminoglycoside-3'-adenylyltransferase"
FT 8157..10244
FT /tag= l
FT /label= pVSI_ORI
FT 10245..11179
FT /tag= m
FT /label= pUC19_ORI
FT 11195..11450
FT /tag= n
FT /label= Nopaline_right_border_fragment
XX
XX WO200068374-A1.
XX
XX 16-NOV-2000.
XX
XX 08-MAY-2000; 2000WO-EP04117.
XX
XX 10-MAY-1999; 99US-0309038.
XX
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Heifetz PB, Patton DA, Levin JZ, Que Q, De Haan PT, Gielens JJL,

```

```

XX WPI: 2001-007392/01.
DR
XX Altering the expression of a viral genome, to obtain cells, plants or
XX animals that are resistant or tolerant to viruses, comprises using
XX sense and antisense RNA fragments that form double-stranded RNA
XX molecule -
PS Claim 14; Page 67-71; 75pp; English.
XX
XX The present invention relates to methods of altering the expression of
XX a viral genome, comprising introducing into a cell, a first DNA sequence
XX capable of expressing a sense RNA fragment of the viral genome and a
XX second DNA sequence capable of expressing an antisense RNA fragment of
XX the viral genome. The sense and antisense RNA fragments are capable of
XX forming double-stranded RNA. The method is useful for rendering cells,
XX CC preferably plant cells, animals or their progenies resistant or tolerant
XX CC to viruses such as tospoviruses, potyviruses, potexviruses,
XX CC tobamoviruses, luteoviruses, cucumoviruses, bromoviruses,
XX CC closteroviruses, tombusviruses and furoviruses.
XX The present sequence is plasmid pZU634, a chimeric gene cassette
XX comprising the coding sequence for coat protein (CP) from potato virus
XX Y (PVY). This plasmid comprises a ubiquitin3 (Ubi3) promoter plus
XX CC leader intron, SMAS promoter, nopaline synthase (NOS) terminator,
XX CC actin2 intron, pVSI ORI, pUC19 ORI and coding sequences for
XX CC phosphomannose isomerase A and bacterial spectinomycin-detoxifying
XX CC enzyme, aminoglycoside-3'-adenylyltransferase. This plasmid is used to
XX CC confer resistance to PVY in tomato.
XX
XX Sequence 11461 BP; 3128 A; 2581 C; 2645 G; 3107 T; 0 other:
SQ

```

```

alignment_scores:
Quality: 145.50 Length: 398
Ratio: 0.841 Gaps: 17
Percent Similarity: 43.467 Percent Identity: 21.357
alignment_block:
US-09-462-846-2 x D02175 ..

```

Align seg 1/1 to: D02175 from: 1 to: 11461

```

10 ProValPheLysGluArgIle.....TTPGlyGI 19
11 : |||||..... |||
12 5380 CCGATCATGCAAAACATCATTAACATGAGCAAAACATGCTGGCGGAG 5429
13
14 19 Y...ThrAlaLeuAlaasp...PheGlyTYrThrIlePro...serGln 33
15 : |||||..... |||||
16 5430 CAAACGGCGGTGACTGAACCTTTATGTAATGGAATAATCCGTCACGCCAGC 5479
17
18 33 rGThrGlyGluCysTrpAlaPheAlaAlaHisGlnAsnGlyGlnSerVal 49
19 : |||||..... |||||
20 5480 CGATGGCGGAGCTGTGG...ATGGCGGCACATCCGAAAGCAGTTGACGA 5526
21
22 50 ValGlnAsnGlyMetTyrLysGlyPheThrIleuSerGluLeuTrrGlnHi 66
23 : |||||..... |||||
24 5527 GTGCAGAAATGCCCGCGAGATATCGTTTCACATCGATGATGATGAAG 5576
25
26 66 sHisArgHis...LeuPheGlyGlnLeuGluLysPArgPhe..... 79
27 : |||||..... |||||
28 5577 TGATTAATGACTCTGCTGCGAGAGCCCTTCCCAACGCTTGGCGAAC 5626
29
30 80 ..ProLeuLeuThrIlySileuAspAlaAspGlnAspLeuSerValGln 95
31 : |||||..... |||||
32 5627 TGGCTTTCTCTTCAAAAGATATATGCGCAGCAGCAGCAGCTCTCCATTGAG 5676
33
34 96 ValHisProAsnAspGluTyrAlaAsnIleHisGluAsnGlyGlnLeuGI 112
35 : |||||..... |||||
36 5677 GTTCATCAACAAC.....AAACACAAATTCTGAATAATCG 5708
37
38 112 Y..... 112
39
40 5709 TTTTGGCAAAAGAAATGCGCGAGGTATCCGATGGATGGCGCGGAGCGTA 5758

```

```

113 .....lysthrGluCysTrpTyrIleIle..... 120
    ||| |||
5759 ACTATAAGATCCCTAACACACAGCCGAGCTGTTTTCGGCGACGCCCT 5808
120 ..... 120
5809 TTCCTGGCATGAACGGCTTCGTGAATTTCCGAGATTGTCCTCTACT 5858
120 ..... 120
5859 CCACCCGGTCCGAGTCACATCCGGCATTCCTCACTTTTACACACCC 5908
121 ..AspCysGlnLysAspAlaGluIleIleTyrGlnHisAsnAlaThrThr 136
5909 CTGATGCGCGAAGCTTTAAGCGAAGCTGTTCCGACGCGTTGAATATGACAG 5958
137 LysGlnGluLeuThrThrMetIle..... 144
    |||||
5959 GGTAAAGAAAATCCCGCGCGCTGGCATTTTAAATCGCCCTCGATAG 6008
145 ..GluArgGlyGlu...TrpAspGlu..... 151
    :::::|||||
6009 CCACAGAGGTGAACCGTCGCAAAAGATTCGTTAATTTCTGAATTTTACC 6058
152 .....LeuLeuArgValLysVal 158
6059 CGAAGACAGCGCTCTCTCCCGCTATTCGTGAATGTGGAAATATG 6108
159 LysProGlnAspPhePheTyrValProSerGlyThrValHisAlaIleG 175
    |||||
6109 AACCTGGCGAAGCGATGTCCTGTCGTGAAGAACCGGACCTTAACT 6158
175 ylyGlyGlyLeuLeuAlaGluThrGlnGlnAsnSerAspThrThrTyr 192
    :::::|||||
6159 GCAAGGGCGCTG...GCGCTGGAGATGATGCAGAACTCCGATTAACGCTG... 6201
192 rGluLeuTyrAspTyrAspArgLysAspAlaGluLysLeuArgGluLeu 208
    |||
6202 .....CTG 6204
209 HisLeuLysLysSerIleGluValIleGluValProSerIleProGluArg 225
    :::::|||||
6205 CGTGGCGGCTGACGCCCTAAATACATTGATATTCGGAACTGGTCCCA 6254
225 gHisThrValHisHisGluGlnIleGluAspLeuLeuThrThrLeuI 242
    :::::|||||
6255 TGTGAATTCGAAAGCAACCGGCTAACCACTGTTGACCCACGCGTGA 6304
242 leGluCysAlaTyr.....PheSerVal 249
    :::::|||||
6305 AACAGAGTGCAGAACTGACTCCCGATTCCAATGATGATTTTGGCTTC 6354
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProP 266
    :::::|||||
6355 TCGCGCATGACCTTAGTGAATAAGAAACACACATTAGCAGACAGATGC 6404
266 leuLeuLeuIleSerValIleGluGluGluGluArgMetIleSerGlyGlu 283
6405 CGCACTTTTGTTCGCGTGAAGCGATGCAAGCTTGTGAAGAGTTCCTC 6454
283 yrfValTyrProPheLysLysGlyAsp..HisMetLeuLeuPro 296
6455 AGCAGTTACAGCTTAACCGGCTGAATCAGCGTTTATTTGCGC 6496
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2001.DAT:DD2174
seq_documentation_block:
ID DD2174 standard: DNA: 12766 BP.
XX
AC DD2174;
XX
DT 28-MAR-2001 (first entry)

```

```

XX Plasmid pZU623 for ZYMV and PRSV resistance in melon.
DE
XX Potyvirus: ZYMV; PRSV: virus resistant; virus tolerant;
KW melon; coat protein; CP; plasmid pZU623; ds.
KW
XX Chimeric - Zucchini yellow mosaic virus.
OS
XX Chimeric - Papaya ringspot virus.
OS
XX Chimeric - Bacteria.
OS
XX Chimeric - Agrobacterium sp.
OS
XX Chimeric - Unidentified.
XX
FH Key
FH Location/Qualifiers
FT 55..1781
FT /tag= a
FT /note= "Ubiquitin3 (UBI3) promoter plus leader
FT intron"
FT 1790..5143
FT /tag= b
FT /note= "PRSV-ZYMV inverted repeat fragment"
FT 1790..2430
FT /tag= c
FT /rpt_type= INVERTED
FT /note= "Corresponds to PRSV coat protein (CP) region"
FT 2511..3111
FT /tag= d
FT /rpt_type= INVERTED
FT /note= "Corresponds to ZYMV CP region"
FT 3245..3686
FT /tag= e
FT /label= Actin2_intronL
FT complement (3822..4422)
FT /tag= f
FT /rpt_type= INVERTED
FT /note= "Corresponds to ZYMV CP region"
FT complement (4503..5143)
FT /tag= g
FT /note= "PRSV CP region"
FT complement (5160..5429)
FT /tag= h
FT /note= "Agrobacterium nopaline synthase terminator"
FT 5498..6669
FT /tag= i
FT /label= SMAS_promoter
FT 6691..7866
FT /tag= j
FT /product= "Phosphomannose isomerase A"
FT 7928..8202
FT /tag= k
FT /label= NOS
FT /note= "Agrobacterium nopaline synthase terminator"
FT 8254..8385
FT /tag= l
FT /label= NOS
FT 8664..9452
FT /tag= m
FT /product= "Bacterial spectinomycin-detoxifying enzyme,
FT aminoglycoside-3'-adenylyltransferase"
FT 9462..11549
FT /tag= n
FT /label= pvs1_ORI
FT 11550..12484
FT /tag= o
FT /label= pUC19_ORI
FT 12500..12755
FT /tag= p
FT /label= Nopaline_right_border_fragment
PD 16-NOV-2000.

```



```

XX 08-MAY-2000; 2000MO-EP04117.
XX
PR 10-MAY-1999; 99US-0309038.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Helfeltz PB, Patton DA, Levin JZ, Que Q, De Haan PT, Gielen JTL;
DR WPI; 2001-007392/01.
XX
PT Altering the expression of a viral genome, to obtain cells, plants or
PT animals that are resistant or tolerant to viruses, comprises using
PT sense and antisense RNA fragments that form double-stranded RNA
PT molecule -
XX
PS Claim 14; Page 62-65; 75pp; English.
XX
CC The present invention relates to methods of altering the expression of
CC a viral genome, comprising introducing into a cell, a first DNA sequence
CC capable of expressing a sense RNA fragment of the viral genome and a
CC second DNA sequence capable of expressing an antisense RNA fragment of
CC the viral genome. The sense and antisense RNA fragments are capable of
CC forming double-stranded RNA. The method is useful for rendering cells,
CC preferably plant cells, animals or their progenies resistant or tolerant
CC to viruses such as tospoviruses, potyviruses, poxviruses,
CC tobamoviruses, luteoviruses, cucumoviruses, bromoviruses,
CC clostoviruses, tombusviruses and furoviruses.
CC The present sequence is plasmid pZU623, a chimeric gene cassette
CC comprising the coding sequence for coat protein (CP) from Zucchini
CC yellow mosaic virus (ZYMV) and Papaya ringspot virus (PRSV). This
CC plasmid also comprises a ubiquitin3 (Ub3) promoter plus leader
CC intron, SMAS promoter, nopaline synthase (NOS) terminator, actin2
CC intron, pVSI ORI, pUC19 ORI and coding sequences for phosphomannose
CC isomerase A and bacterial spectinomycin-detoxifying enzyme,
CC aminoglycoside-3'-adenylyltransferase. This plasmid is used to confer
CC resistance to ZYMV and PRSV in melon.
XX
SO Sequence 12766 BP; 3472 A; 2890 C; 2955 G; 3449 T; 0 other:

alignment_scores:
Quality: 145.50 Length: 398
Ratio: 0.841 Gaps: 17
Percent Similarity: 43.467 Percent Identity: 21.357

alignment_block:
US-09-462-846-2 x D02174 ..

Align seg 1/1 to: D02174 from: 1 to: 12766

10 ProvalphelysgluArgile.....Trrglygl 19
|||||:|||||:|||||:|||||:|||||:|||||:
6685 CCGATCATGCAAAACATTAATCACTGCAAAACATATGCTGGGAG 6734
19 Y...Thralaleualaasp...PheglYtrThrIlePro...Sergina 33
|||||:|||||:|||||:|||||:|||||:|||||:
6735 CAAACGGCGTGTACTGAACCTTATGTATGGAATAATCCGTCACGCCAGC 6784
33 rgtHrGlyglucYstrPralaPhealaIaHIsGlnasnglyGlnserVal 49
|||||:|||||:|||||:|||||:|||||:|||||:
6785 CGATGGCCGAGCTGTGG...ATGGCGGCACATCCAAAGACAGTTCACGA 6831
50 ValGlnasnglyMetYrlysglyPheHrIleuSerGlnleuTrrgluHr 66
|||||:|||||:|||||:|||||:|||||:|||||:
6832 GTGCGAAGAACCGCGGAGATATCGTTTCACGCGGTGATGTAGAGAG 6881
66 sHIsArGHis...LeupheGlyGlnleuGlnlyAspartrPhe..... 79
|||||:|||||:|||||:|||||:|||||:|||||:
6882 TGATTAATGACTGTCTGCGAGAGGCGCTTCCAAACGCTTTGGCGAGAC 6931
80 ..ProleuLeuThrlyslleuAspAlaaspGlnaspLeuSerValGln 95

```

```

6932 TGCCCTTCCTGTTCAAGATATATGCGACAGCAGCCACTTCATCATAG 6981
96 ValHIsProAsnAspGluTrrAlaAsnIleHIsGlnasnglyGlnleu 112
|||||:|||||:|||||:|||||:|||||:|||||:
6982 GTTCATTCCAAC.....AAACACAATTCGAAATCGG 7013
112 Y..... 112
7014 TTTTGCAGAAAGAAATGCCGAGTATCCGATGATGCCCGCAGACGTA 7063
113 .....LysThrGluCysTrpYrIlele..... 120
7064 ACTATAAGATCTTAACACACAGCGAGCTGTTTGGCGTAGACGCT 7113
120 ..... 120
7114 TTCCTTGCATGACGCGTTTCGTGAATTTCCGAGATGTCCTCCTACT 7163
120 ..... 120
7164 CCAGCCGTCGCGAGTGCACATCCGCGATGCTCACTTTTACACAGC 7213
121 ..AspCysGlnLysAspAlaGluIleIleYrGlyHIsAsnAlaTrhr 136
|||||:|||||:|||||:|||||:|||||:|||||:
7214 CTGATGCCGAAGCTTAAAGCAACTGTGCGCAGCTGTGAATATGACG 7263
137 LysGlnGluLeuThrHrMetIle..... 144
7264 GGTGAAGAAATATCCCGCGCGGCGATTTAAATCGGCCCTCGATAG 7313
145 ..GluArgGlyGlu...TrpAspGlu..... 151
|||||:|||||:|||||:|||||:|||||:|||||:
7314 CCAGCAGGAGTGAACCGTGCACAAACGATCTTATTTGATTTTAC 7363
152 .....LeuLeuArgArgValLysVal 158
7364 CGAAGACACGCGTCTGTCTCCCGCTATTCGTGAATGGGGAATGG 7413
159 LysProGlyAspPheHrYrValProSerGlyThrValHIsAlaIleG 175
|||||:|||||:|||||:|||||:|||||:|||||:
7414 AACCTGGCGAAGCGATGTTCTGCTGCGTGAACACCGCAGCTTACCT 7463
175 YLysGlyIleLeuAlaLeuGlnIleHrGlnIleAsnSerAspTrHrThy 192
|||||:|||||:|||||:|||||:|||||:|||||:
7464 GCAAGGCGTG...GCGGTGAAGATGAGCAACCTCGATACAGTG... 7506
192 rgleuTyAspTrYrAspArgLysAspAlaGluGlyLysleuArgGluLeu 208
7507 .....CTG 7509
209 HIsleuLysLysSerIleGluValIleGluValProSerIleProGlu 225
|||||:|||||:|||||:|||||:|||||:|||||:
7510 CGTGGCGGTGACGCTTAATATCATTTATTCGGAACGCTTGCCAA 7559
225 gHIsThrValHIsHIsGlnGlnIleGlnAspLeuLeuThrThrLeu 242
|||||:|||||:|||||:|||||:|||||:|||||:
7560 TGTAATAATGCAAGCCAAACCGGCTAACAGTTGTTGACCACCGGTGA 7609
242 legLucysAlaTrYr.....PheSerVal 249
7610 AACAAAGTGAGAACTGACTTCCGATTCACGTGATGATTTTGCTTC 7659
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnIleLysPro 266
7660 TCGGTGATGAGCTTAAGTAAAGAAACCAACCATTTGACGAGAGTGC 7709
266 eleuLeuIleSerValIleGluGlyGluGlyArgMetIleSerGlyGlu 283
|||||:|||||:|||||:|||||:|||||:|||||:
7710 CGCATTTTGTTCGCTGAGAGCGATGACAGCTTGTAAGAGTTCTC 7759
283 yrValTyProPheLysLysGlyAsp HIsMetleuLeuPro 296
|||||:|||||:|||||:|||||:|||||:|||||:

```

[illegible]

```

225 gHlStHrValHlHlSgluInlIeGlUaSPleuLeuThrThrLeuI 242
      :      :      :      :      :      :      :      :
2886 TGTGAATTCGAAACCAACCGGCTAACCGATTGTGACCCACCGGTGA 2935
      :      :      :      :      :      :      :      :
242 leGlucysAlaTYR.....PheSerVal 249
      :      :      :      :      :      :      :      :
2936 AACAGAGTCGACGAACTGCACTCCGATTCACATGATGATTTGGCTTC 2985
      :      :      :      :      :      :      :      :
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProh 266
      :      :      :      :      :      :      :      :
286 TCGCTGCATACCTTAGTGATAAGAAACCCATTCAGCCAGCAGAGTGC 3035
      :      :      :      :      :      :      :      :
266 eLeuLeuIleSerValIleGlUaSPleuLeuLysGlnGlnLysGlnLys 283
      :      :      :      :      :      :      :      :
3036 CGGCATTTCCTTCGCGTCGACAGCGCATGCAAGCTTGATGGAAGGTTCTC 3085
      :      :      :      :      :      :      :      :
283 yValValTYrProPheLysLysGlyAsp.HisMetLeuLeuPro 296
      :      :      :      :      :      :      :      :
3086 AGCAGTTACAGCTTAAACCGGTCGATCAACGCTTATTGCGC 3127

seq_name: /SID6/gcgdata/geneseq/geneseq/NAL1999.DAT:220087
seq_documentation_block:
ID 220087 standard; DNA; 6103 BP.
XX
XX 220087;
XX
XX 05-JAN-2000 (first entry)
XX
XX Plasmid pCIB9818.
XX
XX Insecticide: HyFLIB; crystal protein; delta-endotoxin; toxin;
KM cryIA(b); maize; transgenic plant; European corn borer;
KM Ostrinia nubilalis; entomocider; crop protection; biological control;
KM pCIB9818; ubiquitin; promoter; mannose phosphate isomerase;
KM pmi gene; selectable marker; ss.
XX
XX Chimeric - Zea mays.
OS Chimeric - cauliflower mosaic virus.
OS Chimeric - Escherichia coli.
XX
XX Key Location/Qualifiers
XX promoter 4..896
      /tag= a
      /note= "maize ubiquitin promoter"
XX misc-feature 2010..3219
      /tag= b
      /note= "PMI gene"
XX intron 3242..3347
      /tag= c
      /note= "inverted PEPC intron 9"
XX terminator 2249..3426
      /tag= d
      /note= "35S terminator"
XX misc-feature 3429..3591
      /tag= e
      /note= "lacZ gene"
XX misc-feature 4050..4910
      /tag= f
      /note= "beta-lactamase (amp) gene"
XX misc-feature 6087..6103
      /tag= g
      /note= "lacZ' gene"
XX
XX W09950293-AL.
XX
XX 07-OCT-1999.
XX
XX 30-MAR-1999; 99WO-EP02175.
XX
XX 01-APR-1998; 98US-0035549.
XX

```

```

PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Desai NM;
XX
XX WPI: 1999-601323/51.
XX
XX Chimeric insecticidal protein comprising cryIB and cryIA(b) portions
XX
XX Disclosure: Page 78-80; 85pp; English.
XX
XX This is the nucleotide sequence of pCIB9818 Ub1/PMI selectable
XX marker plasmid including the mannose phosphate isomerase (PMI)
XX gene under control of the maize ubiquitin promoter. The invention
XX describes the design and construction of a chimeric insecticidal
XX protein, termed HyFLIB (see Y31990), encoded by a synthetic
XX maize-optimized gene (see Z20086), and composed of a cryIB core
XX N-terminal toxin portion and a cryIA(b) C-terminal protoxin
XX portion. When HyFLIB is expressed in transgenic maize,
XX insecticidal activity is observed against European corn borer
XX (Ostrinia nubilalis). Recombinant microbial strains transformed
XX with the HyFLIB gene can be used in endotoxin formulations for
XX the biological control of Lepidopteran pests.
XX
XX Sequence 6103 BP; 1501 A; 1462 C; 1428 G; 1712 T; 0 other;
XX

```

```

alignment_scores:
  Quality: 145.00      Length: 384
  Ratio: 0.868        Gaps: 16
  Percent Similarity: 43.490  Percent Identity: 21.615

```

```

alignment_block:
US-09-462-846-2 x Z20087 ..

```

```

Align seg 1/1 to: Z20087 from: 1 to: 6103

```

```

17 TrpGlyGly...ThrAlaLeuAlaAsp...PheGlyTYrThrIlePro.. 30
   |||||...|||...|||...|||...|||...|||...|||...|||
2046 TGGGCGACGAAACCGCGTTGACGTAACTTATGTATGTAAGAAATCCGTC 2095
   |||||...|||...|||...|||...|||...|||...|||...|||
31 .SerGlnArGThrGlyGlyGlyCysTrpAlaPheAlaAlaHisGlnAsnGly 47
   |||||...|||...|||...|||...|||...|||...|||...|||
2096 CAGCGACGCGATGCGCCGAGCTGTG...ATGGCGCAATCCGAAAGCA 2142
   |||||...|||...|||...|||...|||...|||...|||...|||
47 InSeRValValGlnAsnGlyMetTYrLysGlyPheThrLeuSerGluLeu 63
   :||| :||| :||| :||| :||| :||| :||| :|||
2143 GTTCACGAGTCGACGAATGCCCGCGAGATATCGTTTCACTGCGATGTG 2192
   :||| :||| :||| :||| :||| :||| :||| :|||
64 TrpGlnHisHisArgHis...LeuPheGlyGlnLeuGlnGlyAspArgPh 79
   ||| :||| :||| :||| :||| :||| :||| :|||
2193 ATTGAGAGTGATTAATGCACTCTGCTCGAGAGCGCGTTGCCAAACGCTT 2242
   ||| :||| :||| :||| :||| :||| :||| :|||
79 e.....ProLeuLeuThrLysIleLeuAspAlaAspGlnAsnLeuS 93
   | :||| :||| :||| :||| :||| :||| :|||
2243 TGGCGAAGTGCCTTCCTGCTCAAGATATATGCCACACAGCCACTCT 2292
   ||| :||| :||| :||| :||| :||| :||| :|||
93 eRValGlnValHisProAsnAspGluTYrAlaAsnIleHisGlnAsnGly 109
   ||| :||| :||| :||| :||| :||| :||| :|||
2293 CCATTCAGGTTCATCCAAAC.....AACACAAATTC 2324
   ||| :||| :||| :||| :||| :||| :||| :|||
110 GluLeuGly..... 112
   ||| :||| :||| :||| :||| :||| :||| :|||
2325 GAAATCGGTTTGGCCAAAGAAATGCCGAGTATCCCGATGATGCCGC 2374
   ||| :||| :||| :||| :||| :||| :||| :|||
113 .....LysThrGluCysTrpTYrIleI 120
   ||| :||| :||| :||| :||| :||| :||| :|||
2375 CGAGCGTAAGTATAAGATCCTAACACACAGCCGAGCTGTTTGGGC 2424
   ||| :||| :||| :||| :||| :||| :||| :|||
120 le..... 120
   ||| :||| :||| :||| :||| :||| :||| :|||
2425 TGACGCCCTTCTTCGATGGAACGCGTTTCGTGAATTTCCGAGATGTGC 2474
   ||| :||| :||| :||| :||| :||| :||| :|||

```

120	120
2475	TCCTTAAGCCGGTGGCAGGTGCACATCCGGCATTTGCTCACTTTT	2524
121ASPCyGlnLysAspAlaGluIleIleTyrGlnHisAsnA	134
2525	ACAAACAGCTGATGCCGAAACGTTTAAGGAACGTGTGCCAGCGATTGA	2574
134	LaThrThrLysGlnGluLeuThrThrMetIle.....	144
2575	ATATGACGGGTGAAGAAAAATCCCGCGCGCATTTTAAATCGGCC	2624
145GluArgGlyGlu...TyrAspGlu.....	151
2625	CTCGATAGCCAGCAGCGGTAAACCGTGGCAAAAGATCTTTAATTCTGA	2674
152LeuLeuArgArgy	156
2675	ATTTTACCCGGAAGACAGCGGTCTGTCTCCCGCATTTGCTGAATGG	2724
156	AluLysValLysProGlyAspPhePheTyrValProSerGlyThrValHis	172
2725	TGAATATGAACCTGGCGAAGGATGTCCTGGCGTGAACACCGCAC	2774
173	AlaIleGlyLysGlyIleLeuAlaLeuLeuThrGlnGlnAsnSerAspThr	189
2775	GCATTACCTGCAGAGCGCTG...GCGGTGAAGATGAGCAACCTCGATTAA	2821
189	ThrThrThrGlyLeuTyrAspPyrAspArgLysAspAlaGluGlyLysLeuA	206
2822	CGTG.....	2825
206	ArgLysLeuHisLeuLysLysSerIleGluValIleGluValProSerIle	222
2826CTGCGTGGCGGTCTGACGCGCTAAATACATTGATTTCGGAACTG	2870
223	ProGluArgHisThrValHisHisGlnIleGlnAspLeuLeuThrThr	239
2871	GTTGCCAATGTGAATTTGCAACCCAAACCGGCTAACCGATTGTTCAGCCA	2920
239	ThrThrLeuIleGluGlyAlaTyr.....P	247
2921	GCCGGTGAACAAGGTGCGAACTGACTCCGATTCGACGTGATGATT	2970
247	HisSerValGlyLysTyrPasnLeuSerGlySerAlaSerLeuLysGlnGln	263
2971	TTTCCCTTTCGCTGCATGACTTAATGATTAAGAACAACACATTAGCCAG	3020
264	LysProPheLeuLeuIleSerValIleGluGlnGluGlyArgMetIleSe	280
3021	CAGAGTGCAGCGCATTTTGTCTGCTGCACAAAGCGATGCACAGTTGTGGAA	3070
280	ArgLysGlyLysValTyrProPheLysGlyAspHisMetLeuLeuPro	296
3071	AGGTTTCTACAGATTACACCTTAACCGAGTGAATACCGTTTATTCGCG	3120